

Evidence-based recommendations for designing free-sorting experiments

Simon J. Blanchard¹  · Ishani Banerji²

Published online: 30 September 2015
© Psychonomic Society, Inc. 2015

Abstract The card-sorting task is a flexible research tool that is widely used across many of the subfields of psychology. Yet this same great flexibility requires researchers to make several (seemingly arbitrary) decisions in their designs, such as fixing a sufficient number of objects to sort, setting task requirements, and creating task instructions for participants. In the present research, we provide a systematic empirical investigation of the consequences of typical researcher design choices while administering sorting tasks. Specifically, we studied the effects of seven sorting task design factors by collecting data from over 1,000 online participants assigned to one of 36 sorting tasks, as part of a fractional factorial experimental design. Analyses show the effects of the various researcher decisions on the probability that participants would quit the task, the amount of time spent on the task, the number of piles made, and posttask measures such as satisfaction and depletion. Research design recommendations are provided.

Keywords Free-sorting · Card sorting · Experimental design

In sorting tasks (also known as card sorting), participants are presented with a set of stimuli and asked to sort them into piles. Sorting tasks have a long and rich history in the social sciences and particularly psychology research, and their use has continuously increased over the last few decades, as is

illustrated in Fig. 1, which shows a histogram of a topic search made on the Web of Science for “card sort.”

The areas of application for the sorting task are diverse. In social psychology, Rosenberg, Nelson, and Vivekananthan (1968) asked participants to sort 64 different personality traits (presented on individual pieces of paper) to understand the perceived psychological relatedness of individual traits. More recently, Thomas, Ditzfeld, and Showers (2013) used the task to investigate compartmentalization of the self, so that sorting tasks were the basis for a measure of self-structure: Participants were asked to sort a series of cards that each have a single trait into categories that corresponded to the participants’ various perceived selves. Sorting tasks are also important in assessing how we perceive others. Brewer and Lui (1989) used sorting tasks to investigate how we categorize individuals according to various attributes, and Towler and Schneider (2005) used this task to measure the perceptions of stigmatized groups.

In cognitive psychology, the sorting task has also been used in a variety of contexts. For instance, it is one of the tests that measures executive control in the Delis–Kaplan executive function system (Delis, Kaplan, & Kramer, 2001). In turn, the D-KEFS has been used to assess the effects of various interventions and injuries on executive functioning. For instance, Barbey et al. (2012) used the card-sorting test (in addition to the other D-KEFS tests) to explore the effects of brain lesions on general intelligence and executive control. Within developmental psychology, these types of tasks have been used to compare the cognitive abilities of children with autism, mental retardation, and normal abilities (Shulman, Yirmiya, & Greenbaum, 1995), and more generally to explore the formation of concepts throughout development (Aiken and Williams 1973).

Sorting tasks have also been important in perception research. The importance of haptic versus visual cues for

✉ Simon J. Blanchard
sjb247@georgetown.edu

¹ Department of Marketing, McDonough School of Business, Georgetown University, 37th and O Streets NW, Washington, DC 20057, USA

² Department of Marketing, College of Business, University of Texas San Antonio, One UTSA Circle, San Antonio, TX 78249, USA

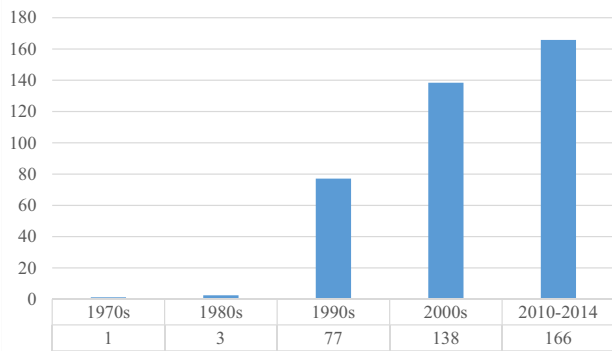


Fig. 1 Web of Science topic search on “Card Sort”: Histogram of average annual number of publication records

cognitive representation was examined using various types of sorting tasks by Klatzky, Lederman, and Reed (1987), who had participants sort objects either by relying solely on touch, solely on their visual perception, or some combination of the two. More recently, a sorting task was used to demonstrate that individuals are fairly inept at perceiving different photographs of the same individual as belonging to one target unless they were previously familiar with the target (Jenkins, White, Van Montfort, & Burton, 2011). Consumer behavior research has also benefited from the use of sorting tasks. For instance, recently Ülkümen, Chakravarti, and Morwitz (2010) used sorting tasks to demonstrate that prior exposure to broad versus narrow categories can have a subsequent spillover effect on information processing and lead consumers to rely on more dimensions in subsequent decision-making tasks. As is illustrated by the use of such different domains, the sorting task is a methodological tool that is relevant to many of the major research questions pursued in cognitive, developmental, perceptual, and social psychology.

Moreover, the sorting task is valuable because it provides participants with a relatively agreeable experience while simultaneously affording researchers a great deal of flexibility. The task has been shown to be both pleasurable to participants (Coxon, 1999) and easy, such that even young children of 3 to 4 years old can participate (Russell & Bullock, 1985). Moreover, participants find sorting tasks to be less tedious than completing pairwise ratings, which would be another method for assessing the perceived relationships among a set of items (Greenberg & McIsaac, 1984), and studies directly comparing the two methods have shown that sorting tasks are a viable alternative to pairwise ratings (Bonebright, 1996).

Because organizing a set of things into a smaller number of groups is perhaps one of the most basic of cognitive and language processes (Coxon, 1999; Quine, 1969), it is no surprise that many variants of the sorting task exist in psychological research. Of relevance to the present research are two variants: closed and free sorting. In a closed sorting task, participants assign objects to a set of predetermined categories (e.g., friend or foe; white, Asian, black) but do so according to their own

perceptions. Such tasks are particularly common in supervised learning studies, whereby a participant learns to correctly categorize a set of objects according to a rule determined by the experimenter (e.g., the famous Wisconsin Card-Sorting Task; Grant & Berg, 1948). The key feature of closed sorting tasks is that a correct sort objectively exists. In contrast, the most unstructured or unconstrained sorting tasks involve asking participants to allocate a set of objects into groups of their own choosing, making as few groups as they wish—a task known as free sorting. Such free-sorting tasks have been used to study unsupervised categorization (Pothos & Chater, 2002) and free grouping (Stringer, 1967). The research discussed below and the studies reported in this article all involve the latter type of task.

When researchers administer a free-sorting task (herein referred to as sorting task, for simplicity), their objective is to study the natural cognitive processing of information that individuals encounter in their lives. As such, researchers must carefully consider the specific aspects of the sorting task to be used, because the various decisions may have a significant impact on the task’s outcomes (i.e., the sorts), thereby influencing the conclusions that can be made about the cognitive processes they are studying. Unfortunately, there is little systematic empirical research that provides guidance on how psychologists should conduct sorting tasks in order to minimize unwanted consequences, such as systematically influencing the obtained data or resulting in depleted or dissatisfied participants. Rather, different studies have provided varying recommendations on the decisions that need to be made when using a sorting task. For instance, previous research has shown that sorting tasks tend to be less taxing than other data collection procedures, such as ratings of pairwise similarity or triadic combinations (Bijmolt & Wedel, 1995), yet recommendations vary with respect to the maximum number of objects that should be accommodated. Whereas Coxon (1999) suggests that it is not uncommon to see tasks with up to 120 objects, the US Department of Health & Human Services (usability.gov) argues that no more than 40 objects should be used, and further suggests that participants should not be required to sort all of the provided objects. The belief is that a greater number of objects will lead to greater fatigue and depletion, and that requiring the use of all objects may obscure the true nature of the results. Similar confusion exists with regard to other decisions, such as whether images should be provided, whether participants should be asked to label the piles (and at what point during the task), whether participants should be allowed to sort objects into more than one pile, and whether participants should be provided with a tutorial that explains how the task should be performed.

In the present study, we provide an empirical investigation of the effects of such researcher-driven free-sorting task design decisions on a variety of outcomes, including participants’ experiences of the task. To do so, we randomly administered one of 36 different sorting tasks to a large sample of

online panel participants, generated following a fractional factorial design that systematically varied the design of each sorting task according to typical researcher decisions (Addelman, 1962; see also Collins, Dziak, & Li, 2009). We then analyzed the resulting data to show which factors affected the sorting task outcomes and participants' experiences. Doing so allowed us to provide guidance as to the best practices and potential pitfalls in the design of sorting task data collection. We made the choice to focus on free sorting (as opposed to other types of sorting tasks, such as closed sorts, or Q-sort) because there are more free parameters for researchers, and thus an empirically based article that outlines the effect of researcher decisions would be more valuable.

The remainder of the article is structured as follows. In the next section, we describe the sorting task configuration decisions that have to be made by the researcher, along with (a) typical configuration options. Then we identify measures that characterize the sorting task data collected and describe a post-task survey that allowed us to examine how participants perceived the experience. In the empirical section, we present the generated experimental design and descriptive statistics of the collected data. Using a series of random-effects linear models, we report the effects of the manipulated researcher decisions on a variety of outcomes. We conclude in the General Discussion by formulating precise recommendations.

Researcher decisions and sorting task structures

Although sorting tasks are particularly easy for participants to complete, generating one involves many moving pieces that may appear to be ad-hoc to a researcher. To investigate how such arbitrary decisions influence the outcomes of the sorting task, we create an experimental design that systematically varies seven different decisions that a researcher may face when implementing a sorting task. In the present section, we focus on describing how prior research has dealt with the decisions regarding these design elements. Doing so allowed us to determine the typical research design characteristics that would be the basis of our experimental design.

Decision 1: Pretask tutorials

Although online interfaces that use a simple drag-and-drop interface (e.g., cardsorting.net) can be quite intuitive, participants do not always fully understand the task at hand. For instance, Lickel et al. (2000) provided a pretask example that read as follows:

In a sorting task, your job is simply to put things together into “types” as you see fit. For example, consider the following list of foods. [The page showed a list of foods, numbered 1–8, and three boxes into which the foods had been sorted, according to vegetables, fruits, and sweets.]

As you can see, the foods that are of the same type have been put in the same box. You may have seen a different way of grouping these foods together. Your way of grouping them would have been just as valid as the one we provided here. (p.228)

Despite requiring participants to read this tutorial, Lickel et al. (2000) had to eliminate 36 out of 199 participants (18%) because they either did not complete the task or repeatedly assigned a group to multiple piles. Such high incomplete rates bring into question the necessity of administering clear pretask instructions.

In the present set of studies, we chose to offer either no prior example or a short video that illustrated how to drag objects in and out of piles, how to label the piles (if applicable), and so forth. The tutorial explained the task by demonstrating the sorting of musical instruments. We chose musical instruments because they are easily sorted into different piles, but at the same time were quite different from the objects that participants would be sorting in the actual task (food items and groups; see Decision 3). Different versions of the tutorial were created to reflect the nature of the task that the participant would be completing. For instance, if participants would be labeling piles during the task (Decision 4), the video demonstrated labels being changed. However, if participants would not be labeling their piles during the actual sorting task, the video did not show labels. Regardless of the version (i.e., food or group items) presented, all of the tutorial videos demonstrated musical instruments being sorted.

Decision 2: Number of objects

The researcher must decide on the number of objects that participants will be asked to sort. Whereas common sizes for offline sorting tasks are rarely recommended to exceed 20–30 objects, online interfaces featuring easy drag-and-drop functionalities (e.g., cardsorting.net) may be able to accommodate more objects without making the task unmanageable. How should a researcher determine the maximum number of objects to include? One option would be to refer to previous studies. However, researchers' perceptions of what participants would find manageable seem to have varied greatly. For instance, Ross and Murphy (1999) provided 45 objects, Lickel et al. (2000) provided 40 objects, and Isen and Daubman (1984) provided 36 objects. In every case, the number of objects presented to participants was argued to be an appropriately manageable number that would not affect the category structure recovered. Yet, Rosenberg et al. (1968) provided 64 objects, and Brewer and Lui (1989) provided 140 in their studies, while arguing that the set of objects represented a broad enough array to be representative of the sample space, without being too taxing for participants. This raises the question: How many is too many? In our experimental

design, we investigated whether increasing the number of objects to be sorted had downstream consequences on the enjoyment of the task, the time spent on the task, and the number of categories generated. We varied the number of objects to be sorted from 20, to 40, to 60, bearing in mind that these encapsulate a range of the values most typically considered.

Decision 3: Types of objects to be sorted

One of the key features of the sorting task is that it allows for the study of a wide variety of objects. The sorting task has been used to study the categorization of concrete objects such as trees, foods, and music (songs), as well as more abstract items such as groups and emotions. Previous research has documented that several differences may emerge in a categorization process that involves abstract objects over more concrete objects. For instance, concrete objects are more likely to be categorized on their basic-level perceptual features first, before semantic memory is searched for higher-level categorizations (Jolicœur, Gluck, & Kosslyn, 1984; Murphy & Smith, 1982). In contrast, abstract concepts (i.e., entities that are neither purely physical nor spatially constrained) may be more difficult to retrieve from memory than concrete concepts (Barsalou & Wiemer-Hastings, 2005). Thus, it may be that asking participants to sort difficult abstract objects (e.g., the group types from Lickel et al. 2000) may deplete participants more than asking them to sort a similar number of more concrete and easily sortable objects (e.g., the common food objects from Ross & Murphy, 1999).

Furthermore, for concrete objects, one must decide whether images should be used. Previous research has shown that the addition of images in sorting tasks can have a significant impact on the labels provided. For instance, Jørgensen (1995) found that 73% of the labels used when text-only objects are sorted were related to perceptual features (e.g., attributes). However, this percentage dropped to 20% when images were also provided. Similarly, Rorissa and Iyer (2008) showed that using images led to a significant increase in the use of superordinate category labels. Thus, providing images may lead participants to categorize at a less basic level, and lead participants to see objects as possibly fitting multiple piles simultaneously.

In our investigation, we used two types of objects that differed substantially. First, we used the food objects from Ross and Murphy (1999), since they represent a very concrete set of objects. Using food objects allowed us to have a set of objects associated with known categories, and for which pictures could also be displayed (see, e.g., Jolicœur et al. 1984, who used 24 pictures of similar food objects). We also used the list of groups in Lickel et al. (2000), which were the basis for identifying a taxonomy of “group types”—arguably a much more abstract set of objects than food items. The lists of food objects and groups used in these experiments are

presented in Tables 1 and 2, respectively. Thus, this researcher decision involved three options in our analyses: food objects, food objects with images (75 pixels × 75 pixels), and group names. Note that these differ from the objects (i.e., musical instruments) that were sorted in the tutorial videos presented to some participants.

Decision 4: If (and when) to ask for pile labels

The researcher must decide on whether he or she wishes participants to label the piles that they make, and if so, whether this should occur during or after participants have submitted the piles. In some cases, participants may not need to generate the labels for each pile. For example, Hunt and Seta (1984) asked participants to sort words into predetermined piles, and the labels for each pile were provided to participants. However, in many instances researchers may desire participants to generate the labels for each pile themselves. Without doubt, analyzing the text data from participant labels can be burdensome, but prior research has demonstrated that it can be useful to a researcher who wishes to understand the basic reasons why participants put objects in the same piles (Gopnik & Meltzoff, 1992).

Indeed, asking participants to label piles during the task is a common practice (cf. Coxon, 1999) that is assumed not to impact the cognitive process, such that it may be possible to measure category structures during the task in the same way that nonconscious goals can be measured outside of awareness (Carlson, Tanner, Meloy, & Russo, 2014). Yet, asking for labels during the task may have a significant effect on the cognitive processes. We know from previous research that participants prefer to use rule-based categorization when prompted for justification (Smith & Sloman, 1994) and tend to report categories at the most basic level (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). This raises the possibility that asking participants to label and sort simultaneously leads them to sort objects into piles that are easy for them to justify. In the hope of minimizing such effects, an alternative strategy is to ask participants to label the piles after the sorting is finished (e.g., Blanchard, 2011). In our experimental investigation, we systematically varied whether the labels were requested from participants during the task (as they grouped objects into piles), after the task (once their sorts were submitted), or not at all.

Decision 5: Providing criteria for the sorts

The researcher must decide on instructions to provide to participants regarding the basis on which they should sort the items. Sometimes these instructions can be quite vague, due to the nature of the study. For example, Gopnik and Meltzoff (1987) asked young infants to sort objects. Thus, the only instruction provided to these

Table 1 Food object sorting stimuli

Item	Added to Set ^a	Tasks ^b	Item	Added to Set ^a	Tasks ^b	Item	Added to Set ^a	Tasks ^b
Carrots	No	20,40,60	Apple	no	20,40,60	Chicken	no	20,40,60
Lettuce	no	20,40,60	Orange	no	20,40,60	Hamburger	no	20,40,60
Corn	no	20,40,60	Pineapple	no	40,60	Salmon	no	40,60
Potato	no	40,60	Banana	no	40,60	Lobster	no	40,60
Onions	no	40,60	Watermelon	no	60	Steak	no	40,60
Broccoli	no	40,60	Strawberry	yes	60	Pork	no	
Cucumber	yes	60	Kiwi	yes	60	Tofu	yes	
Tomato	yes	60	Spaghetti	no	20,40,60	Shrimp	yes	
Soda	no	20,40,60	Bread	no	20,40,60	Milk	no	20,40,60
Water	no	20,40,60	Muffin	no	20,40,60	Eggs	no	20,40,60
Coffee	yes	40,60	Rice	no	40,60	Yogurt	no	40,60
Tea	yes	60	Pizza	no	40,60	Butter	no	40,60
			Cereal	no	40,60	Cheese	no	40,60
Cake	no	20,40,60	Pancakes	no	40,60	Margarine	yes	60
Pie	no	20,40,60	Crackers	no	60	Greek yogurt	yes	60
Cookies	no	40,60	Bagel	no	60			
Doughnuts	no	40,60	Oatmeal	no	60	Potato chips	no	20,40,60
Ice cream	no	60	Hot roll	yes	60	Pretzels	no	20,40,60
Popsicles	yes	60	Bagel	yes	60	Nuts	no	20,40,60
Cupcakes	yes	60				Chocolate bar	no	40,60
						Popcorn	no	40,60
						Granola bar	no	60
						Gummy bears	yes	60

^a Objects added (yes) indicate food objects that were not part of the original Ross and Murphy (1999) stimuli. ^b Numbers indicate the sets in which the item is included. For instance, “40, 60” indicates that the item is assigned when the number of objects to be sorted is determined to be either 40 or 60, but not 20.

participants was that they should play with the objects or “fix them up” (p.1526). In other cases, the researcher may be interested in a particular set of outcomes and instruct participants accordingly. For instance, in a study in which the main dependent measures of interest were participants’ response time and error rates, the instructions provided to participants emphasized that they should sort as quickly and accurately as possible (Klatzky et al. 1987). However, the most common set of instructions involves asking participants to sort the objects into piles such that they “go together,” on the basis of perceived similarity between the objects (Coxon, 1999). This is not surprising, given that psychology has long documented a strong relationship between categorization and judgments of similarity (Nosofsky, 1986; Rips, 1989).

An interesting distinction becomes necessary when one considers that the relationship between similarity and dissimilarity (e.g., psychological distance) may be more complicated than one expects (Tversky & Gati, 1978), particularly in the context of sorting task data (Burton, 1975). Specifically, previous research has indicated that judgments of similarity lead participants to focus on the common features between stimuli

more than judgments of dissimilarity. Consequently, depending on the nature of the stimuli, judgments of similarity and dissimilarity need not always be complementary (Tversky & Gati, 1978). In the present investigation, we chose contrasting instructions; one set of participants were instructed to sort on the basis of perceptions of similarity (“create piles such that objects in the same piles are similar in some way”), whereas other participants received instructions that explicitly requested dissimilarity (“create piles such that objects in different piles are different in some way”).

Decision 6: Type of sorting task

When researchers are interested in understanding the dominant associations between objects, the common single card-sorting task is typically administered, in which participants are allowed to place each item into only one pile (Coxon, 1999). That is, participants cannot place the same item into more than one pile. This procedure forces participants to assign objects to the piles that they see as having the strongest connection to the objects, and thus requires them to make trade-offs during the sorting.

Table 2 Group-type sorting stimuli

Item	Added to Set ^a	Tasks ^b	Item	Added to Set ^a	Tasks ^b
Members of a family	no	20,40,60	People at a bus stop	no	20,40,60
Two people in a romantic relationship	no	20,40,60	People in line at a bank	no	20,40,60
Friends who do things together	no	20,40,60	People in the audience at a movie	no	20,40,60
Members of a local street gang	no	20,40,60	People who live in same neighborhood	no	20,40,60
Members of an orchestra	no	40,60	Students at a university	no	20,40,60
Members of a sports team	no	40,60	Students enrolled in a class	no	20,40,60
Members of a rock band	no	40,60	People living in a retirement home	no	40,60
Married couples	yes	40,60	People who enjoy classic music	no	40,60
Members of a military unit	yes	60	People attending an athletic contest	no	40,60
Medical interns training together	yes	60	Plumbers	no	40,60
Prison cellmates	yes	60	Teachers	no	40,60
Hostages in an abduction	yes	60	People who work in the same factory	no	40,60
			Members of a labor union	no	40,60
Airline flight crew	no	20,40,60	Members of the same political party	no	60
A company committee designing a new product	no	20,40,60	People on the same airplane	yes	60
Coworkers assigned to a project	no	20,40,60	People at a coffee shop	yes	60
Members of a jury	no	20,40,60	People treated at the same hospital	yes	60
Members of the cast of a play	no	20,40,60	Guests of the same hotel	yes	60
Members of a student campus committee	no	20,40,60	Holocaust survivors	yes	60
Members of a local environmental organization	no	40,60	Coin collectors	yes	60
Members of a support group	no	40,60			
Students studying for an exam	no	40,60	Citizens of Poland	no	20,40,60
Employees of a local restaurant	no	40,60	Citizens of America	no	20,40,60
Members of a university social club	no	40,60	Women	no	20,40,60
People having dinner together on a Saturday night	no	40,60	Blacks	no	20,40,60
Roommates	no	60	Jews	no	40,60
Members of a fraternity	yes	60	Doctors	no	40,60
Members of a grant review committee	yes	60	Lawyers	yes	40,60
Members of the same congregation	yes	60	Men	yes	60
People who play street basketball	yes	60	Arabs	yes	60
Co-founders of a startup	yes	60	Vegetarians	yes	60

^a Objects added (yes) indicate group names that were not part of the original Lickel et al. (2000) stimuli. ^b Numbers indicate the sets in which the item is included. For instance, “40, 60” indicates that the item is assigned when the number of objects to be sorted is determined to be either 40 or 60, but not 20

Yet, in many settings, participants may perceive the objects as fitting into multiple piles, and a researcher may instead allow participants to sort the objects into more than one pile if they so desire (Blanchard, 2011; Blanchard & DeSarbo, 2013). For instance, consider a sorting task in which participants sort pictures (faces) of individuals. A participant may wish to sort objects into three distinct piles: a “woman” pile, a “man” pile, and an “Asian” pile. If a single card-sorting task is used, the participant will be forced to make a trade-off that may not reflect the true nature of his or her perceptions. In this case, the participant may sort the picture of an Asian woman into a “woman” summary pile. It would be incorrect to conclude from this behavior that the participant views an Asian woman as a woman but not as Asian. Simply put, the use of single card sorting tasks may lead to an increase in false

negatives when objects are seen by participants as potentially belonging to multiple categories.

These trade-offs are avoided in the multiple card-sorting task, which allow a participant to sort objects into multiple piles; thus, the participant could place the picture of the Asian woman into both the “woman” and “Asian” piles. As such, conclusions based on non-assignments of objects to piles (i.e., saying that an object is not seen as representative of the pile) are only really possible using a multiple card-sorting task. It is thus no surprise that Lickel et al. (2000) stated that they had to discard a significant number of participants who placed a group into multiple piles, consistent with prior research that had concluded that single-sort data (in which participants are only allowed to sort objects into one pile) “may not adequately represent the psychological

categories and dimensions of a stimulus domain” (Rosenberg & Park Kim, 1975, p.497). In our experimentation, we systematically varied whether the participants were given the opportunity to sort objects into multiple piles (multiple card-sorting task; Blanchard, 2011; Blanchard & DeSarbo, 2013) or the (traditional) single card-sorting task.

Decision 7: Requiring participants to use all of the cards at least once

There may be situations in which a participant is not familiar with all of the objects provided, or in which the researcher expects participants to skip over some of the cards included in the sorting portion of the task. In those cases, a researcher may wish to ensure that all participants use all of the cards at least once, a common requirement in paper-based sorting tasks (Coxon, 1999; Rosenberg & Park Kim 1975). Yet, in other situations it would be preferable not to require participants to use all cards at least once. For instance, when sorting a large number of objects, participants may be overwhelmed and decide not to use all of the objects. It may also be the case that some other aspect of the study design unrelated to the sorting task itself could make it difficult to require participants to use all of the cards. For example, Gopnik and Meltzoff (1987) asked infants who were less than 2 years old to sort objects, and it would have been too difficult to require that these participants use all of the objects for their sorts. Alternatively, the sorting task may involve some objects that are unfamiliar to the participants. In these conditions, the researcher can adopt one of two strategies: instruct participants to put all of the unknown objects into an “unknown” pile, to be labeled as such, or ask participants to sort only the objects that they are familiar with. In the present set of tasks, we systematically varied whether or not participants were required to use all of the cards presented.

In summary, in the present research we systematically manipulated seven different decisions that may be faced by a researcher who is considering using a sorting task. In the next section, we describe the study in more detail and list the various outcomes of interest that these decisions might impact.

Free-sorting experimental design

Ideally, we would like to investigate the effect of researcher decisions using a full factorial design, such that for each possible combination of researcher decisions, we would collect a different set of randomly assigned panel participants. Doing so would allow us to investigate complex interactions between researcher decisions and their effect on the task outcomes. However, the number of tasks, and consequently the numbers of participants, would be prohibitive. Specifically, we would have to generate 432 ($3^3 2^4$) different sorting tasks, each with

many randomly assigned participants. Instead, we used a fractional factorial design to study the main effect of each factor (researcher decision) on various dependent measures (Addelman, 1962; see also Collins et al. 2009), in keeping with previously published methodological research in the psychometric literature that had investigated the effects of data-generating mechanisms (e.g., Blanchard & DeSarbo, 2013; DeSarbo, 1982; Helsen & Green, 1991). The result was a reduced set of 36 tasks with orthogonal factors that would allow us to investigate the effects of researcher decisions on task outcomes. The complete design is presented in Table 3.

Participants

We used Amazon Mechanical Turk (mTurk) for our studies, because it provides a platform on which we could study normal populations (Paolacci & Chandler, 2014). We posted a request for participants for a “consumer perceptions study” in exchange for \$0.50. In the description, participants were told that the task took 12 min on average, but that variance in completion times could occur. We requested 720 participants so that we would have 20 per task, which we allocated evenly (sequentially) using Qualtrics’ quota functions. Our sample was 34.99 years old ($SD = 12.34$) on average, 51.4% female, 75.4% Caucasian, 7.3% African American, 6.8% Hispanic, 8.1% Asian, and 2.5% other racial backgrounds.

Materials and procedure

Once a participant had clicked the survey link on mTurk, each participant was randomly assigned to one of the 36 task configurations via an initial Qualtrics survey whose sole purpose was to randomly assign participants. If the participant was assigned to a task design for which pretask examples were to be provided (*Decision 1: Using pretask tutorials*), Qualtrics displayed a short video (30 s) in which musical instruments were sorted, along with a set of instructions that followed those described in Lickel et al. (2000). The video differed depending on whether participants were assigned to a single or a multiple card-sorting task (*Decision 6: Type of sorting task*), and whether they were asked to label their piles during the task (*Decision 4: If and when to ask for pile labels*). For instance, a multiple card-sorting task in which participants were to label the piles during the task (not before or after) was preceded by a video that included images and the following subtitles:

Drag and drop the objects into piles. Feel free to move the objects from one pile to the next based on your own perceptions. You can even put the objects into multiple piles. In this example, violin, harp and clarinet are in more than one pile. Note you can also remove objects from their piles. Or take them out completely.

Table 3 Sorting task Monte Carlo design: Average task outcome results, per task

Task	Objects	Labeling	Card Type	Nb. Cards per Item	Using All Cards Once	Instructions	Pretask Tutorials	% of Workers Who Dropped	Nb. Piles	Nb. Cards Unused	Nb Used More Than Once	Time Spent Sorting	Average Participant Congruence
1	20	After	Food Objects	One	Required	Similarity	Without	16.67%	6.64	0.00	0.48	171.96	.539
2	20	After	Food Objects	Multiple	Not Required	Dissimilarity	With	25.00%	5.71	0.00	0.48	138.29	.591
3	20	During	Food Objects+ Pictures	One	Not Required	Dissimilarity	With	32.14%	5.95	0.00	0.00	151.42	.583
4	40	During	Food Objects	Multiple	Required	Dissimilarity	With	25.00%	7.62	0.00	2.19	353.81	.547
5	60	During	Groups	One	Not Required	Dissimilarity	Without	57.14%	9.75	0.00	0.00	500.75	.508
6	20	After	Food Objects	One	Not Required	Dissimilarity	Without	39.29%	5.82	0.00	0.00	132.18	.589
7	40	After	Groups	One	Not Required	Similarity	With	50.00%	10.93	1.20	0.26	321.13	.438
8	20	None	Groups	Multiple	Required	Dissimilarity	Without	32.14%	7.05	0.00	0.00	313.68	.490
9	60	During	Groups	One	Required	Similarity	Without	32.14%	9.95	0.00	0.00	664.05	.510
10	60	None	Food Objects	One	Not Required	Similarity	With	10.00%	9.52	0.00	0.00	337.19	.575
11	60	After	Food Objects+ Pictures	One	Not Required	Similarity	With	35.71%	7.78	0.00	0.00	446.33	.582
12	20	During	Food Objects+ Pictures	Multiple	Not Required	Similarity	Without	32.14%	7.32	0.00	0.84	285.95	.539
13	40	None	Food Objects+ Pictures	One	Not Required	Dissimilarity	Without	17.86%	8.17	0.00	0.00	248.22	.551
14	40	None	Food Objects+ Pictures	Multiple	Not Required	Dissimilarity	With	25.00%	9.43	0.86	1.57	327.65	.533
15	40	After	Groups	Multiple	Required	Dissimilarity	Without	17.86%	8.87	0.00	0.09	423.26	.452
16	60	After	Food Objects+ Pictures	Multiple	Required	Dissimilarity	Without	39.29%	8.47	0.00	5.65	575.81	.586
17	40	After	Groups	One	Required	Dissimilarity	With	21.43%	8.91	0.00	0.67	277.00	.458
18	20	None	Groups	Multiple	Not Required	Similarity	Without	35.71%	7.89	0.00	0.00	204.06	.458
19	40	During	Food Objects	One	Not Required	Dissimilarity	Without	32.14%	7.26	0.00	0.00	332.61	.568
20	60	After	Food Objects+ Pictures	One	Required	Dissimilarity	With	28.57%	9.80	0.00	0.00	285.55	.508
21	20	During	Food Objects+ Pictures	Multiple	Required	Similarity	With	36.67%	6.68	0.00	2.26	241.89	.551
22	40	None	Food Objects+ Pictures	One	Required	Similarity	Without	21.43%	8.09	0.00	0.48	286.82	.543
23	20	After	Food Objects	Multiple	Required	Similarity	With	25.00%	5.67	0.00	0.00	129.33	.599
24	40	None	Food Objects+ Pictures	Multiple	Required	Similarity	With	32.14%	7.42	1.36	1.84	334.21	.565
25	40	During	Food Objects	Multiple	Not Required	Similarity	Without	21.43%	8.59	0.05	0.05	320.55	.554
26	60	During	Groups	Multiple	Not Required	Dissimilarity	With	32.14%	9.42	4.32	1.32	681.06	.507
27	20	None	Groups	One	Required	Dissimilarity	With	39.29%	8.29	0.00	0.12	221.76	.420
28	40	After	Groups	Multiple	Not Required	Similarity	Without	39.29%	9.71	0.00	0.58	327.18	.489
29	60	During	Groups	Multiple	Required	Similarity	With	57.14%	7.75	0.00	0.00	576.25	.501
30	60	None	Food Objects	One	Required	Similarity	Without	21.43%	11.91	0.00	0.00	401.41	.498
31	20	None	Groups	One	Not Required	Similarity	With	25.00%	5.81	0.00	0.00	202.14	.524
32	20	During	Food Objects+ Pictures	One	Required	Dissimilarity	Without	21.43%	6.36	0.00	0.00	190.05	.613
33	60	After	Food Objects+ Pictures	Multiple	Not Required	Similarity	Without	13.33%	7.23	0.00	2.85	457.12	.600
34	60	None	Food Objects	Multiple	Not Required	Dissimilarity	With	17.86%	8.35	0.00	0.48	348.59	.599
35	40	During	Food Objects	One	Required	Similarity	With	14.29%	7.25	0.00	0.00	249.04	.566
36	60	None	Food Objects	Multiple	Required	Dissimilarity	Without	25.00%	10.19	0.00	0.38	484.05	.548

The four videos and instructions are available on the first author's homepage or upon e-mail request.

After viewing the video, participants proceeded to the sorting task interface. The participants who were assigned to a task design that did not provide pretask examples proceeded to the sorting task interface without watching the videos. We used Cardsorting.net, which affords researchers the ability to present a varying number of objects to be sorted (*Decision 2: Number of objects*), to customize instructions (*Decision 5: Providing criteria for the sorts*), to use either a single or a multiple card-sorting task (*Decision 6: Type of sorting task*), and the option to ask participants to label the piles during or after the task (*Decision 4: If and when to ask for pile labels*). The text instructions used are presented in [Appendix A](#).

When participants submitted their completed sort, the interface inspected the data to determine whether the task parameters had been met and prompted participants if this was not the case. Specifically, participants whose task required them to sort all of the objects at least once (*Decision 7: Requiring participants to use all of the cards at least once*) were told: "You have some unused objects on the menu to the left. This study requires you to use all the objects at least once. Please put all of them once on the board before proceeding!" Below the message, a list displayed all of the unsorted objects. Participants not required to sort all of the objects at least once were also provided the list of the unsorted objects, but the message was as follows: "You have some unused objects. Are you sure you want to finalize your study? Here are some of the objects that you have not used. Please drag them into the desk before continuing." However, they were allowed to continue and submit their sorts if they so desired. After submitting their sorts, participants whose task required them to label the piles after the task (*Decision 4: If and when to ask for pile labels*) were asked to do so. Afterward, all participants proceeded to the same posttask survey, which contained some of our key dependent measures.

Measures from the sorting task

For each administered task, we obtained a number of dependent measures that are the focus of our descriptive analyses. These measures (shown in [Tables 3 and 4](#)) include six task outcomes that were calculated from the sorting task data itself, and a set of postsorting task survey outcomes.

Dropout percentages The first immediate consequence of the researcher's decisions is whether participants will choose to carefully complete the task. In online panels, many participants fail to submit the task, but instead decide to abandon it without completing the task. As such, it would be important to determine whether any decisions made by the researcher (e.g., the number of objects to be sorted, the lack of pretask

examples, etc.) may lead to increased attrition of participants from the study.

Although we requested 720 participants from mTurk, all the participants who began the study did not complete it. Using Qualtrics quotas to assign participants to our 36 conditions, we were able to see that mTurk had assigned 1,018 participants to our 36 tasks. This resulted in 738 participants who completed one of the tasks in its entirety. We found that 17 participants misunderstood the task instructions (e.g., failed to drag the objects into piles, resulting in one pile per object), leaving 721 participants for our analyses. Looking at the quota records from Qualtrics, we were able to identify the number of participants who dropped out of each of the 36 tasks, thus making it possible to compute and compare the dropout percentages associated with various researcher decisions.

Time spent A common concern is that sorting tasks involving a large number of items may take an inordinate amount of time for participants, and that the resulting data may be of poor quality. Therefore, we measured the total time (in seconds) spent on the sorting portion of the task.

Number of piles made The objects chosen for sorting (food objects and group names) had been shown in previous research to correspond to a defined set of categories (seven for the food objects: vegetables, fruits, drinks, desserts, carbohydrates, meats, dairies, and snacks; four for the group names: intimacy groups, task groups, categorical groups, and loose associations). However, the various researcher decisions might influence whether participants sorted the objects into similar numbers of categories. In order to determine the effect of researcher decisions on the number of categories, we computed the number of piles made by each participant.

Number of cards unused We expected that if participants did not like the task (the objects were too hard to sort, too uninteresting, too many, etc.), then a greater number of cards would remain unused. This measure, of course, was only available for the tasks in which participants were not forced to use all of the cards.

Using cards more than once Whereas we expected that participants would only use the same card in multiple piles when a strong conflict arose in their perceptions, it is possible that other researcher decisions might also lead to this behavior. Consequently, we computed the frequency with which participants used the same card more than once for those participants who were assigned to the tasks in which this was possible.

Congruence between sorts In the context of free sorting, analysis does not generally involve a "true answer," such that we cannot meaningfully test whether a given manipulation increased or decreased task performance. We can, however,

Table 4 Sorting task Monte Carlo design: Average participant perception measures, per task

Task	Objects	Labeling	Card Type	Nb. Cards per Item	Using All Cards Once	Instructions	Pretask Tutorials	Task: Satisfaction	Task: Enjoy	Task: Difficulty	Task: Challenging	Task: Effort	Cards Simil	Cards Varié	Time Solv.	Time Unsol
1	20	After	Food Objects	One	Required	Similarity	Without	5.32	5.17	2.52	2.97	3.44	5.48	5.40	31.81	118.65
2	20	After	Food Objects	Multiple	Not Required	Dissimilarity	With	5.38	5.35	1.95	2.50	3.05	5.33	5.00	43.18	152.47
3	20	During	Food Objects+ Images	One	Not Required	Dissimilarity	With	5.26	5.28	2.42	2.98	3.37	4.53	5.32	28.34	119.61
4	40	During	Food Objects	Multiple	Required	Dissimilarity	With	5.24	5.29	2.76	3.74	4.71	4.81	5.14	61.62	140.55
5	60	During	Groups	One	Not Required	Dissimilarity	Without	4.67	4.42	3.83	4.21	4.58	4.08	5.50	23.39	123.13
6	20	After	Food Objects	One	Not Required	Dissimilarity	Without	4.82	4.88	2.65	3.41	4.18	4.76	4.82	44.67	143.40
7	40	After	Groups	One	Not Required	Similarity	With	4.47	4.56	4.33	4.14	4.67	4.60	5.33	34.37	140.14
8	20	None	Groups	Multiple	Required	Dissimilarity	Without	4.74	4.53	3.58	3.95	4.42	4.42	4.95	51.76	128.17
9	60	During	Groups	One	Required	Similarity	Without	4.74	4.79	4.32	4.53	4.74	4.89	4.53	29.82	101.39
10	60	None	Food Objects	One	Not Required	Similarity	With	4.63	4.68	2.93	3.55	4.30	4.89	5.85	23.57	95.60
11	60	After	Food Objects+ Images	One	Not Required	Similarity	With	4.78	4.50	2.50	3.28	4.00	4.33	4.83	39.75	156.20
12	20	During	Food Objects+ Images	Multiple	Not Required	Similarity	Without	4.68	4.72	2.79	3.18	4.26	5.00	5.11	46.53	127.69
13	40	None	Food Objects+ Images	One	Not Required	Dissimilarity	Without	5.04	4.67	2.83	3.28	3.74	5.13	5.78	42.75	106.37
14	40	None	Food Objects+ Images	Multiple	Not Required	Dissimilarity	With	4.90	4.14	3.38	3.86	4.33	4.86	5.19	32.93	132.29
15	40	After	Groups	Multiple	Required	Dissimilarity	Without	4.65	4.41	3.48	3.98	4.39	4.43	5.00	23.98	119.66
16	60	After	Food Objects+ Images	Multiple	Required	Dissimilarity	Without	4.94	4.47	3.24	3.92	4.59	4.53	5.65	46.83	135.57
17	40	After	Groups	One	Required	Dissimilarity	With	4.82	4.65	3.95	4.39	4.82	4.55	4.91	54.80	111.95
18	20	None	Groups	Multiple	Not Required	Similarity	Without	4.33	4.54	3.61	3.98	4.11	4.56	4.83	49.31	81.88
19	40	During	Food Objects	One	Not Required	Dissimilarity	Without	5.32	4.82	2.74	3.28	3.84	4.21	5.47	78.13	137.20
20	60	After	Food Objects+ Images	One	Required	Dissimilarity	With	4.90	4.88	3.10	3.60	4.15	4.90	5.75	35.54	106.59
21	20	During	Food Objects+ Images	Multiple	Required	Similarity	With	5.21	5.32	2.42	3.52	4.00	5.11	4.58	54.76	84.38
22	40	None	Food Objects+ Images	One	Required	Similarity	Without	5.23	5.11	2.77	3.11	3.59	4.82	5.41	60.75	154.50
23	20	After	Food Objects	Multiple	Required	Similarity	With	5.48	5.29	2.90	3.39	3.95	4.95	5.48	39.63	119.51
24	40	None	Food Objects+ Images	Multiple	Required	Similarity	With	4.89	4.86	2.53	3.55	4.58	5.16	5.32	29.77	188.39
25	40	During	Food Objects	Multiple	Not Required	Similarity	Without	4.91	5.05	3.23	3.84	4.45	5.00	5.55	42.52	121.56
26	60	During	Groups	Multiple	Not Required	Dissimilarity	With	4.32	4.12	4.37	4.88	5.32	4.68	4.84	41.39	103.38
27	20	None	Groups	One	Required	Dissimilarity	With	4.35	4.96	3.88	4.22	4.47	4.88	4.82	26.86	82.91
28	40	After	Groups	Multiple	Not Required	Similarity	Without	4.65	4.35	3.59	4.03	4.47	4.94	4.41	42.68	90.87
29	60	During	Groups	Multiple	Required	Similarity	With	4.75	4.39	4.42	5.13	5.83	5.42	5.58	30.10	98.09
30	60	None	Food Objects	One	Required	Similarity	Without	4.59	4.55	2.64	3.44	4.00	5.23	5.77	42.04	151.58
31	20	None	Groups	One	Not Required	Similarity	With	4.38	4.76	3.48	4.02	4.57	5.19	4.29	42.81	109.40
32	20	During	Food Objects+ Images	One	Required	Dissimilarity	Without	5.09	4.92	2.45	2.94	3.45	4.95	5.45	29.30	133.70
33	60	After	Food Objects+ Images	Multiple	Not Required	Similarity	Without	5.08	5.01	2.35	3.42	4.04	4.81	5.15	55.54	154.73
34	60	None	Food Objects	Multiple	Not Required	Dissimilarity	With	5.17	4.87	2.65	3.15	3.52	4.26	5.43	50.38	122.46
35	40	During	Food Objects	One	Required	Similarity	With	4.58	4.28	2.08	2.68	3.42	4.92	5.33	43.50	91.75
36	60	None	Food Objects	Multiple	Required	Dissimilarity	Without	5.14	4.84	2.67	3.33	4.00	4.71	5.62	34.94	187.43

quantify how, within a given task, the sorts differed from one another as a measure of consistency between participants.

To do so, we first converted each participant's sort matrix (dimensions: # piles \times # objects) into individual matrices of co-occurrences (pairwise counts; dimensions: # objects \times # objects). Doing this transformation allowed us to compare sorts across individuals (within a task), regardless of the number of piles they made. Given any two participants' pairwise count matrix, we could follow Brusco and Cradit (2005), who suggested vectorizing the lower triangular and performing a Procrustes rotation to eventually obtain the congruence coefficient between two sets of data: a number bounded between 0 and 1 that indicates the similarity of different configurations. Using this measure we can quantify, for each task, how congruent were the sorts made by participants as a measure of consistency. By calculating the congruence between each pair of participants who completed the same task and by taking the average, we obtained the average pairwise participant congruence.

Postsorting task survey measures Once participants finished the sorting task, they proceeded to a short survey. The survey was meant to capture participant perceptions of the task. In particular, it measured satisfaction with the task, enjoyment, difficulty, and effort required. In addition, participants were asked to rate the extents to which they considered the objects to be similar to each other (perceptions of similarity) and to differ from one another (perceptions of variety) (Mogilner, Rudnick, & Iyengar, 2008). Our questions and scales are presented in Appendix B.

After answering these questions, participants were told they would be asked to solve six anagrams. Specifically, they were told: "You'll now be doing a short task in which you'll be asked to solve anagrams. This is not a test. Work on them for as long as you want, and when you want to stop, click continue." This first screen included, in randomized order, three solvable anagrams: CHNUL (LUNCH), WETLO (TOWEL), and ENEGR (GREEN). The second screen included three unsolvable anagrams: ONECI, RTEAN, and ACELO. Unsolvable tasks (such as anagrams) have often been used as indicators of depletion (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Vohs & Baumeister, 2004). Thus, the time spent on attempting the unsolvable anagrams served as an indicator of the extent to which the various researcher decisions might lead to participant depletion.

Results

After administering and collecting the data for our samples, we summarized the data for our analyses. The descriptive statistics, means, and percentages for each of our conditions are broken down by task in Table 3 (measures from the sorting task) and Table 4 (measures from the posttask survey). For all

analyses of our measures (except for dropout rates), we analyzed the effect of the task structures via mixed-effect linear models in which the task number was a random effect and all of the researcher decisions were fixed effects. The dropout rates were analyzed via linear regression, since we had only one observation per task. We note that when the researcher decision involved three alternatives (number of objects, requesting pile labels, and type of cards), variables were coded in terms of two binary indicators with a middle category set to zero. Doing so allowed us to consider any coefficient as the simple effect of the researcher's decision, conditional on the other decisions being at their most common levels. Following the default levels provided in Tables 5 and 6, the standard selected was a single card-sorting task with 40 objects.

When the decision involved only two alternatives (task type, using all cards at all once, instruction types, and providing pretask sorting tutorials), the binary indicators were mean-centered so that simple effects could be evaluated at mean levels. The coefficients reported for these factors can be interpreted as half of the difference in means between the two decision alternatives. These results are summarized in Tables 5 and 6.

Effect of providing pretask tutorials

We found that whereas providing a tutorial to participants had no effect on their experiences with the study (all p s $>$.10), it did help participants sort faster (approximately 80 s), because they got familiar with the task more quickly ($\beta = -41.23$; $Z = -2.69$, $p < .01$). Another noteworthy consequence is that a tutorial helped participants figure out how they could remove objects from their piles via drag and drop, such that participants not required to allocate all objects to at least one pile were slightly more likely to leave a few objects unused ($\beta = 0.76$; $Z = 1.78$, $p < .10$), a difference on average of 1.52 objects left unused. Providing a presorting example is thus advisable if participants are not under tight time pressure.

Effect of the number of objects

Although the number of objects to be sorted is often difficult to determine, we found that the decision can have a significant (mostly negative) effect when it becomes as large as 60. The number of objects to be sorted was also found to have a very significant effect on the amount of time that participants spent sorting the objects. For instance, going from 20 to 40 objects increased the average time spent by nearly 2 min ($\beta = 116.10$; $Z = 6.24$, $p < .01$), and going from 40 to 60 objects increased it by close to 3 min ($\beta = 163.77$; $Z = 8.72$, $p < .01$). However, the number of objects had no effect on the dropout rates. $\beta = -1.87$; $Z = -4.23$, $p < .01$) Thus, a larger set size meant that participants took longer to sort the objects, but it did not lead them to abandon the task entirely.

Table 5 Analysis results: Effect of task design factors on sorting tasks

Task Factor	% Participants Dropped			Number of Piles Made			Number of Cards Unused			Number of Cards Used More than Once			Time Spent Sorting			Average Congruence		
	Coef.	<i>t</i>	<i>z</i>	Coef.	<i>z</i>		Coef.	<i>z</i>		Coef.	<i>z</i>		Coef.	<i>z</i>		Coef.	<i>t</i>	
Intercept	0.26	6.01	***	8.04	17.19	***	1.09	2.06		0.47	0.79		315.17	15.96	***	.574	42.61	***
Pretask Tutorial (default: Not provided)	0.01	0.23		-0.38	-1.05		0.76	1.78	*	0.02	0.05		-41.23	-2.69	***	0.00	0.04	
Objects (default: 40)	-0.02	-0.45		-1.87	-4.23	***	-0.77	-1.61		-0.05	-0.09		-116.10	-6.24	***	-0.02	-1.71	*
Task Type (default: Single card)	0.04	0.82		0.77	1.74	*	-0.07	-0.15		0.81	1.46		163.77	8.72	***	0.02	1.85	*
Multiple Cards ^b	0.01	0.29		-0.23	-0.64		0.65	1.71	*	-	-		61.07	3.99	***	0.01	0.86	
Pile Labeling (default: During)	-0.07	-1.73	*	0.65	1.44		-1.10	-2.28	*	-0.63	-1.07		-71.72	-3.80	***	-0.02	-1.48	
Instructions (default: Dissimilarity)	-0.03	-0.78		0.02	0.04		-0.88	-1.79	*	0.18	0.31		-75.12	-3.95	***	-0.01	-0.64	
Card Categories (default: Food)	-0.01	-0.17		0.01	0.03		-0.29	-0.71		-0.57	-1.08		-0.32	-0.02		0.00	-0.23	
Using All Cards Once (default: Not required)	0.14	3.26	***	0.78	1.73	*	0.91	1.86	*	-0.08	-0.13		113.48	5.99	***	-0.08	-6.71	***
Food with Pictures ^a	0.05	1.25		-0.21	-0.48		0.01	0.03		1.99	3.51	***	34.42	1.88	**	0.00	0.03	
Required ^b	-0.02	-0.57		0.22	0.60		-	-		0.53	1.18		23.04	1.5		-0.02	-1.73	*
Number of obs.	36			722			358			358			716			36		
Number of tasks	N/A			36			18			18			36			N/A		

^a three-category variables are coded as two indicators. Simple effects of the other factors are evaluated at the default level. The coefficient represents the effect on the dependent variable of moving from the default to the alternative. ^b binary variable was mean-centered, so that the simple effects of the other variables are evaluated at the mean value. The coefficient, once doubled, represents the effect on the dependent variable of moving from the default to the alternative. * $p < .10$. ** $p < .05$. *** $p < .01$

Table 6 Analysis results: Effect of task design factors on post-sorting task survey measures

Task Factor	Task Satisfaction		Task Enjoyment		Task Difficulty		Task Effort		Perceived Cards Similarity		Perceived Variety in Cards		Time Spent on Solvable Anagrams		Time Spent on Unsolvable Anagrams	
	Coef.	z	Coef.	z	Coef.	z	Coef.	z	Coef.	z	Coef.	z	Coef.	z	Coef.	z
Intercept	5.06	44.34 ***	4.84	42.51 ***	2.74	20.64 ***	4.06	28.2 ***	4.83	38.08 ***	5.33	38.62 ***	49.58	10.23 ***	124.00	9.95 ***
Pretask Tutorial (default: Not provided)	-0.03	-0.37	0.04	0.41	0.03	0.32	0.15	1.30	0.05	0.48	-0.05	-0.43	-3.41	-0.91	-10.16	-1.05
Objects (default: 40)	0.05	0.45	0.30	2.81 ***	-0.24	-1.92 *	-0.32	-2.32 **	0.16	1.3	0.14	1.06	-5.16	-1.13	-9.23	-0.78
Task Type (default: Single card)	-0.09	-0.81	-0.03	-0.31	0.11	0.85	0.16	1.17	-0.07	-0.58	0.25	1.89 *	-7.31	-1.59	0.61	0.05
Pile Labeling (default: During)	0.09	0.98	0.01	0.07	0.02	0.17	0.25	2.21 **	0.00	0.02	-0.06	-0.55	3.55	0.95	8.41	0.87
Instructions (default: Dissimilarity)	-0.08	-0.75	-0.07	-0.61	-0.06	-0.49	-0.18	-1.28	0.06	0.48	0.03	0.23	-2.33	-0.5	13.28	1.12
Card Categories (default: Food)	0.08	0.75	0.03	0.31	-0.12	-0.92	-0.19	-1.36	0.02	0.19	-0.13	-0.99	-1.72	-0.37	14.33	1.19
Using All Cards Once (default: Not required)	-0.12	-1.41	0.02	0.25	-0.04	-0.38	0.09	0.77	0.28	2.79 ***	-0.09	-0.82	-0.70	-0.19	-6.08	-0.63
Food with Pictures ^a	-0.49	-4.46 ***	-0.38	-3.45 ***	1.25	9.82 ***	0.80	5.80 ***	-0.17	-1.42	-0.55	-4.18 ***	-5.88	-1.27	-23.53	-1.96 **
Required ^b	-0.04	-0.35	-0.08	-0.76	0.08	0.69	0.11	0.83	-0.03	-0.23	-0.12	-0.91	-2.03	-0.45	2.90	0.25
Number of obs.	0.10	1.09	0.10	1.08	-0.01	-0.10	0.07	0.65	0.14	1.40	0.14	1.33	-2.17	-0.58	3.96	0.41
Number of tasks	722	36	722	36	722	36	722	36	722	36	722	36	721	36	721	36

^a three-category variables are coded as two indicators. Simple effects of the other factors are evaluated at the default level. The coefficient represents the effect on the dependent variable of moving from the default to the alternative. ^b binary variable was mean-centered, so that the simple effects of the other variables are evaluated at the mean value. The coefficient, once doubled, represents the effect on the dependent variable of moving from the default to the alternative. * $p < .10$. ** $p < .05$. *** $p < .01$

The number of objects also had an important effect on the numbers of piles made by participants. Our objects were chosen such that they fit fairly obvious categories (seven food objects and four group types). Yet, we found that changing the number of objects influenced the structures made. First, decreasing the number of options from 40 to 20 decreased the number of piles made by almost two, on average (and somewhat decreased the average congruence between participants ($\beta = -0.02$; $t = -1.71$, $p < .10$). The effect was significant but less strong for a reduction from 60 to 40 in both the numbers of piles ($\beta = 0.77$; $Z = 1.74$, $p < .10$) and the average congruence between participants ($\beta = 0.02$; $t = 1.85$, $p < .10$). We note that a separate analysis on the rates (time spent divided by number of items to be sorted) revealed no difference in rates between having to sort 20 or 40 items (2.39 s per card; $\beta = 2.40$; $Z = 0.35$, $p = .73$), but a marginal difference between 40 and 60 items: 11.82 s per card slower with 60 items ($\beta = 11.82$; $Z = 1.71$, $p = .08$). This suggests that a significant increase in complexity occurs when participants go from 40 to 60 items, such that the rate at which each item is sorted becomes longer.

The number of objects to be sorted had a strong effect on participants' experiences with the task. We found that, unsurprisingly, participants enjoyed the task more when they were sorting fewer objects ($\beta = 0.30$; $Z = 2.81$, $p < .01$). Moreover, they found it to be somewhat less difficult ($\beta = -0.24$; $Z = -1.92$, $p < .10$) and effortful ($\beta = -0.32$; $Z = -2.32$, $p < .05$) than when 40 or more objects had to be sorted. However, the number of objects did not have any significant effect on participants' satisfaction with the task or their depletion (measured in terms of the time spent on the solvable and unsolvable anagrams). In addition, we observed no effect on the perceived similarity among the objects, and only a marginal effect on the perceived variety among the objects ($\beta = 0.25$; $Z = 1.89$, $p < .10$).

Effect of the type of object to be sorted

Participants were administered cards representing either types of groups or types of food. The provided types of food were presented either with just the names of the food objects or with images along with the names. We first compared the participants who were asked to sort group names to the participants who were asked to sort food objects (both with and without pictures). The researcher's choice of which objects to administer had the most significant impact on the sorting process. Namely, participants spent nearly 2 min longer sorting groups than sorting food objects ($\beta = 113.48$; $Z = 5.99$, $p < .01$). In addition, participants provided significantly longer labels ($\beta = 7.89$; $Z = 11.50$, $p < .01$) and sorted the objects into a moderately higher number of piles ($\beta = 0.78$; $Z = 1.73$, $p = .10$) when they were sorting group names. This was particularly surprising, given that the group names were expected to result in fewer piles (four, according to Lickel et al.

2000) than the food objects (seven, from Ross & Murphy, 1999). Thus, the group names seem to be associated with higher effort on the part of participants, which we will see later was borne out in their perceptions of the task. This result is further supported by the finding that the sorts made with the groups were a lot less congruent between participants than those made with the food items ($\beta = -0.08$; $t = -6.71$, $p < .01$). It is not surprising, then, that the participants who were sorting group names were also more likely to leave cards unused when allowed to do so ($\beta = 0.91$; $Z = 1.86$, $p = .10$), and were 14% more likely to abandon the task ($\beta = 0.140$; $t(35) = 3.26$, $p < .01$) than participants sorting food objects.

We also found that the decision to offer groups instead of food objects to sort had a significant impact on participants' experiences of the task. Specifically, participants sorting groups saw less variety in the objects they were sorting ($\beta = -0.55$; $Z = -4.18$, $p < .01$), were less satisfied [$\beta = -0.49$; $t(35) = -4.46$, $p < .01$], enjoyed the task less ($\beta = -0.38$; $Z = -3.45$, $p < .01$), and felt it was more difficult ($\beta = 1.25$; $Z = 9.82$, $p < .01$) and effortful ($\beta = 0.80$; $Z = 5.80$, $p < .01$) than those sorting food objects. These participants were depleted to the point that they spent 23 s less on solving the unsolvable anagrams ($\beta = -23.53$; $Z = -1.96$, $p < .05$). Overall, it is fair to say that participants had a significantly less positive experience when sorting the group objects instead of the food objects.

Adding images to the food objects did not have a large impact on the sorting process or on participants' perceptions of the task. We found that participants sorting food objects with images spent about 20 s on average more ($\beta = 34.42$; $Z = 1.88$, $p < .05$), perhaps because the cards were made bigger and thus were slightly more difficult to sort into piles. Moreover, the number of objects sorted into more than one pile substantially increased from 0.49 cards to 2.48 cards ($\beta = 1.99$; $Z = 3.51$, $p < .01$). Thus, whereas it took participants longer to complete the sorts, adding images to the cards afforded them the opportunity to identify instances in which the same item might fit into more than one category. No significant effects emerged on participants' perceptions of the task, in terms of either the perceived similarity of the objects or the perceived variety in the provided set.

Effect of asking participants to label the piles

The researcher may decide to ask participants to label the piles while sorting the objects, or after the sorting has been completed, or not at all. This decision influenced the time it took participants to complete the sorting, the likelihood that they would not use all of the cards, and the dropout percentage. Not surprisingly, participants spent approximately 1 min more on the task when they were asked to label the piles during the sorting (no labeling, $\beta = -71.72$; $Z = -3.80$, $p < .01$; labeling during the task, $\beta = -75.12$; $Z = -3.95$, $p < .01$). In addition, we found that asking participants to label their piles during the

task led them to be somewhat more likely to leave some cards unused (no labeling, $\beta = -1.10$; $Z = -2.28$, $p < .10$; labeling during the task, $\beta = -0.88$; $Z = -0.88$, $p < .10$). Finally, we found a marginal effect on dropout rates, such that participants who realized they would have to label the piles while sorting were more likely to drop the task altogether ($\beta = -0.074$; $t(35) = -1.725$, $p < .10$). The decision to ask for labels (and when) had no effect on how participants perceived the sorting task. That is, we found no effects on perceived effort, difficulty, enjoyment, perceived variety and similarity of the objects, and satisfaction with the task. In addition, the decision had no effect on participant depletion.

Effect of the criterion (similarity vs. dissimilarity)

We found that asking participants to sort on the basis of similarity (as opposed to dissimilarity) had very little effect on the sorting processes (all $ps > .10$). It also had no effect on the participants' experience, other than leading them to perceive the objects as being more similar under the similarity instructions than under the dissimilarity instructions ($\beta = 0.28$; $Z = 2.79$, $p < .01$). Although effectively a manipulation check, this result helps us have additional confidence in the ability of participants to follow instructions and to focus on the aspect of the objects that is most important to the research goal, as emphasized by the instructions.

Effect of the type of sorting task (single card vs. multiple cards)

Providing participants an option to sort the objects into multiple piles meant that they took, on average, 2 min longer to complete the task ($\beta = 61.07$; $Z = 3.99$, $p < .01$). It also meant that when they were not required to use all of the cards at least once, they were slightly more likely to leave some cards unused ($\beta = 0.65$; $Z = 1.71$, $p < .10$). In general, participants found the multiple sorting to be more effortful ($\beta = 0.25$; $Z = 2.21$, $p < .05$), but not necessarily more difficult ($\beta = 0.02$; $Z = .17$, n.s.) or any less enjoyable ($\beta = 0.01$; $Z = .07$, n.s.). Of note, it also did not affect perceptions of the variety of the objects to be sorted or their perceived similarity, and the effort was not so high as to deplete participants to the point that they would spend less time on the follow-up anagram tasks (all $ps > .10$).

Effect of requiring participants to use the cards at least once

Requiring participants to use all of the cards at least once had no significant effect on most of our measures, including drop rates, the numbers of piles made, and the amount of time spent sorting, and had no effect on participants' experiences with the task. We did, however, find that it decreased the average congruence between participants ($\beta = -0.02$; $t = -1.73$, $p < .10$).

General discussion

Free sorting is a data collection methodology used by a large number of psychologists from just as many disciplines. Both online panel data and computerized platforms for free sorting are becoming more readily available, thus making it possible to conduct these experiments even more easily than before. Yet researchers have many decisions to make regarding how the task should be presented to the participants.

In the present study, we showed that such arbitrary decisions can have important consequences on how the task is performed and how it is perceived by participants. Specifically, we generated a set of 36 different online sorting tasks that allowed us to investigate the effect of various sorting task configurations on the outcomes and on participants' experiences. On the basis of our analyses and the 16 dependent variables studied, we provide the following recommendations:

Researchers should keep the number of objects manageable, but there does not seem to be a need to severely constrain the number of objects Whereas there is no doubt that participants prefer tasks with fewer objects (say 20), we found no evidence that they cannot properly follow instructions that require them to sort 40, or even 60, objects. As the number of objects increases from 20 to 40, the biggest impact seems to be on completion time, since the task is perceived to be more effortful and somewhat less enjoyable. However, the jump from 40 to 60 objects seemed to have little effect beyond making the participants feel there was little variety in the options to be sorted. Whereas we have no doubt that participants could complete tasks with a larger number of objects, we found that 40–60 seems to be a reasonable upper bound, on the basis of our empirical evidence.

Researchers should be mindful of what they are asking participants to sort Our results suggest that participants are willing to engage in sorting tasks with a large number of objects. Yet they are significantly less satisfied with the task if they do not enjoy the topic at hand because the objects are perceived to be too similar, or just not interesting. We found that using groups instead of food objects led to an increase of 14% in dropout rates, nearly 2 min longer spent sorting (an equivalent number of objects), and that participants were significantly depleted from the sorting task that they were significantly less likely to work hard on subsequent tasks.

There seem to be few reasons not to allow participants to sort objects into multiple piles Our results suggest that, when given the option, participants were likely to assign objects to multiple piles only when their perceptions dictated it. We did find that this increased the amount of time participants spend on the sort slightly, and the sort was perceived as being more effortful, but it did not affect their perception of task difficulty or their satisfaction with the task. Furthermore, allowing

multiple sorts avoids having to eliminate participants who do sort objects into multiple piles because this naturally fits their perception of the objects. Given that in some previous research data had to be eliminated due to participants frequently sorting the same item into more than one category (e.g., Lickel et al. 2000), our results suggest that this may not be necessary. Allowing participants to do so may reflect their natural cognitive processes and may result in less data attrition.

If researchers want labels for the piles, they should ask participants to do so after they have submitted their sorts as complete Our data suggest that asking participants to name the piles during the task leads to an increase in dropout rates of an average of 5%, a greater number of cards left unused by participants, and a significantly longer time spent sorting.

If researchers are trying to increase the frequency at which participants assign objects to multiple piles, they should consider providing pictures along with the objects' names Doing so allows participants to visualize the objects, and tends to lead to a greater number of objects being used more than once in the sorts.

Requiring participants to use all of the cards has little effect on task and satisfaction measures If you suspect that participants will be familiar with the majority of the objects to be sorted, then requiring them to sort all objects is more likely to result in fully complete data. We found that doing so had nearly no effect on the process or on how participants experienced the task.

Sorting interfaces are sufficiently intuitive; expansive instructions may not be necessary We found that participants did follow the instructions, and that providing participants with videos that illustrated the features of the sorting interface (e.g., adding/removing objects from piles, removing them from the desk, and labeling) did not have much of an impact other than to speed up the sorting process once participants got started. As a researcher, keep your instructions simple, explain clearly to participants how you want them to sort the objects, and they will likely follow your instructions.

In general, we suggest that researchers who have been wary of using sorting tasks thus far in their research should consider using online interfaces (e.g., Qualtrics and cardsorting.net) that provide a great deal of flexibility and feasibility to the researcher, while simultaneously providing participants with a reasonably pleasant and interesting task. Moreover, online interfaces allow rapid data collection, thus making it possible to collect large amounts of data cheaply and quickly via online panels such as mTurk. Our results in Fig. 1 suggest that the use of sorting tasks is indeed becoming more common, and the increased availability of such data-gathering platforms would only help stimulate research.

Through this initial experimental investigation, we have established that researchers should carefully consider the decisions they make when designing free-sorting experiments. Our investigation explored seven common types of such decisions, but they are by no means exhaustive of all possible decisions that researchers may make, and our results need to be interpreted in context. For instance, we only studied sorting tasks administered via an online interface, and as such our findings should be interpreted carefully when they are considered for designing sorting tasks that involve physical paper cards or actual objects. Furthermore, since we did not have a priori hypotheses as to which factors would interact, our fractional experimental design did not allow us to meaningfully interpret second-order (or greater) interactions. Our results should be interpreted cautiously, in that the effects were evaluated at the default levels indicated. It would be particularly interesting for researchers to develop and test hypotheses as to which of the tested factors interact via new experimental design configurations.

Finally, it is important to note that the present analyses do not investigate the effect of the design choices on latent category structures (i.e., the composition of the resulting piles). We chose not to do so, because this would introduce several potential sources of bias. First, although many procedures exist that can help analyze sorting data, the choice of which to use depends on the researcher's objective and is somewhat arbitrary. To illustrate, we considered using aggregate multidimensional scaling and cluster analyses (e.g., Lickel et al. 2000) and Robinson matrices (Ross & Murphy, 1999). Yet such aggregate or sample-level procedures mask respondent heterogeneity, creating risks of aggregation bias, which has been shown to be likely in the context of categorization and similarity (cf. Blanchard, 2011; Brusco & Cradit, 2005). Also, whereas other methodologies do capture participant heterogeneity and are tailored for the analysis of sorting data (e.g., Blanchard et al. 2012; Blanchard, Aloise, & DeSarbo, 2012; Brusco & Steinley, 2013), such procedures are each only appropriate with one type of sorting task (either single or multiple card-sorting tasks). None could be applied to our entire design, which creates complications regarding assessments of fits and comparing structures across tasks. Given these risks and our desire to minimize bias, we thus purposely chose to decouple data collection and statistical inference regarding latent category structures. To help stimulate research in this area, we note that researchers may obtain our entire experimental data files by either contacting the first author or by downloading our compressed data archive of the datasets.¹

Finally, our findings suggest that it may be worthwhile to revisit existing findings while varying researcher decisions. Doing so may bring interesting nuances into existing yet robust findings, and thus contribute to growing psychological theory using free-sorting data.

¹ The archive is available at <http://www.cardsorting.net/tutorials/BRMfiles.zip>.

Appendix A

Table 7 Sorting task instructions, by condition

Card Type	Instructions	Single Card-Sorting Task	Multiple Cards-Sorting Task
Groups	Dissimilarity	<p>On this page, you will find a list of various group names which we would like you to sort into piles such that each pile represents a “type” of group.</p> <p>Drag all the group names into piles, such that the groups in different piles represent different “types” of groups. In other words, the groups in different piles should be dissimilar to each other in some way.</p> <p>You may put as many or as few of the groups together as a “type.” There is no right or wrong way of sorting the group names into piles.</p> <p>To remove a group from a pile that you’ve already made, simply drag the group to the left-side menu or to a different pile.</p>	<p>On this page, you will find a list of various group names which we would like you to sort into piles such that each pile represents a “type” of group.</p> <p>Drag all the group names into piles, such that the groups in different piles represent different “types” of groups. In other words, the groups in different piles should be dissimilar to each other in some way.</p> <p>You may put as many or as few of the groups together as a “type.” You can also use the same group in more than one pile if that’s what makes sense to you. There is no right or wrong way of sorting the group names into piles.</p> <p>To remove a group from a pile that you’ve already made, simply drag the group to the left-side menu or to a different pile.</p>
	Similarity	<p>On this page, you will find a list of various group names which we would like you to sort into piles such that each pile represents a “type” of group.</p> <p>Drag all the group names into piles, such that the groups in the same pile represent a single “type” of group. In other words, the groups in a pile should be similar to each other in some way.</p> <p>You may put as many or as few of the groups together as a “type.”</p> <p>There is no right or wrong way of sorting the group names into piles.</p> <p>To remove a group from a pile that you’ve already made, simply drag the group to the left-side menu or to a different pile.</p>	<p>On this page, you will find a list of various group names which we would like you to sort into piles such that each pile represents a “type” of group.</p> <p>Drag all the group names into piles, such that the groups in the same pile represent a single “type” of group. In other words, the groups in a pile should be similar to each other in some way.</p> <p>You may put as many or as few of the groups together as a “type.” You can also use the same food item in more than one pile if that’s what makes sense to you. There is no right or wrong way of sorting the group names into piles.</p> <p>To remove a group from a pile that you’ve already made, simply drag the group to the left-side menu or to a different pile.</p>
Food objects	Dissimilarity	<p>On this page, you will find a list of various food objects which we would like you to sort into piles such that each pile represents a “type” of food.</p> <p>Drag all the food objects into piles, such that the food objects in different piles represent different “types” of food. In other words, the food objects in different piles should be dissimilar to each other in some way.</p> <p>You may put as many or as few of the food objects together as a “type.”</p> <p>There is no right or wrong way of sorting the group names into piles.</p>	<p>On this page, you will find a list of various food objects which we would like you to sort into piles such that each pile represents a “type” of food.</p> <p>Drag all the food objects into piles, such that the food objects in different piles represent different “types” of food. In other words, the food objects in different piles should be dissimilar to each other in some way.</p> <p>You may put as many or as few of the food objects together as a “type.” You can also use the same food item in more than one pile if that’s what makes sense to you.</p> <p>There is no right or wrong way of sorting the group names into piles.</p>
	Similarity	<p>On this page, you will find a list of various food objects which we would like you to sort into piles such that each pile represents a “type” of food.</p> <p>Drag all the food objects into piles, such that the food objects in the same pile represent a single “type” of food. In other words, the food objects in a pile should be similar to each other in some way.</p> <p>You may put as many or as few of the food objects together as a “type.”</p> <p>There is no right or wrong way of sorting the group names into piles.</p>	<p>On this page, you will find a list of various food objects which we would like you to sort into piles such that each pile represents a “type” of food.</p> <p>Drag all the food objects into piles, such that the food objects in the same pile represent a single “type” of food. In other words, the food objects in a pile should be similar to each other in some way.</p> <p>You may put as many or as few of the food objects together as a “type.” You can also use the same food item in more than one pile if that’s what makes sense to you.</p> <p>There is no right or wrong way of sorting the group names into piles.</p>

Appendix B

Table 8 Postsorting task survey measures

We're now going to ask a few questions about the task you completed earlier.

- (Satisfaction) How satisfying did you find the task?
Very Dissatisfied (1) — Very Satisfied (7)
- (Enjoyment 1) How pleasant did you find the task?
Very Unpleasant (1) — Very Pleasant (7)
- (Enjoyment 2) How interesting or tedious did you find the task?
Very Tedious (1) — Very Interesting (7)
- (Enjoyment 3) How fun or boring did you find the task?
Very Boring (1) — Very Fun (7)
- (Difficulty) How easy or difficult did you find the task?
Very Easy (1) — Very Difficult (7)
- (Effort) How effortful did you find the task?
Very Effortless (1) — Very Effortful (7)
- Think back to the list of objects you were asked to sort into piles.
- (Perceived Similarity) How similar were the objects to one another?
Very Dissimilar (1) — Very Similar (7)
- (Perceived Variety) How much variety was there in the objects?
Very Little Variety (1) — A Lot of Variety (7)

Note: Enjoyment measures had a reliability coefficient (Cronbach's alpha) of .89

References

- Adelman, S. (1962). Orthogonal main-effect plans for asymmetrical factorial experiments. *Technometrics*, *4*, 21–46.
- Aiken, L. S., & Williams, T. M. (1973). A developmental study of schematic concept formation. *Developmental Psychology*, *8*, 162–167.
- Barbey, A. K., Colom, R., Solomon, J., Krueger, F., Forbes, C., & Grafman, J. (2012). An integrative architecture for general intelligence and executive function revealed by lesion mapping. *Brain*, *135*, 1154–1164.
- Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129–163). Cambridge, UK: Cambridge University Press.
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*, 1252–1265.
- Bijmolt, T. H., & Wedel, M. (1995). The effects of alternative methods of collecting similarity data for multidimensional scaling. *International Journal of Research in Marketing*, *12*, 363–371.
- Blanchard, S. (2011). *A methodology for identifying unobserved categories when consumers assign brands to multiple categories* (Unpublished doctoral dissertation). Philadelphia, PA: Pennsylvania State University.
- Blanchard, S. J., Aloise, D., & DeSarbo, W. S. (2012). Heterogeneous p-median for categorization based clustering. *Psychometrika*, *77*, 741–762. doi:10.1007/s11336-012-9283-3
- Blanchard, S. J., & DeSarbo, W. S. (2013). A new zero-inflated negative binomial methodology for latent category identification. *Psychometrika*, *78*, 322–340.
- Bonebright, T. L. (1996). An investigation of data collection methods for auditory stimuli: Paired comparisons versus a computer sorting task. *Behavior Research Methods, Instruments, & Computers*, *28*, 275–278. doi:10.3758/BF03204780
- Brewer, M. B., & Lui, L. N. (1989). The primacy of age and sex in the structure of person categories. *Social Cognition*, *7*, 262–274.
- Brusco, M. J., & Cradit, J. D. (2005). ConPar: A method for identifying groups of concordant subject proximity matrices for subsequent multidimensional scaling analyses. *Journal of Mathematical Psychology*, *49*, 142–154. doi:10.1016/j.jmp.2004.11.004
- Brusco, M. J., & Steinley, D. (2013). Model selection for minimum-diameter partitioning. *British Journal of Mathematical and Statistical Psychology*, *67*, 471–495.
- Burton, M. L. (1975). Dissimilarity measures for unconstrained sorting data. *Multivariate Behavioral Research*, *10*, 409–423.
- Carlson, K. A., Tanner, R. J., Meloy, M. G., & Russo, J. E. (2014). Catching nonconscious goals in the act of decision making. *Organizational Behavior and Human Decision Processes*, *123*, 65–76.
- Collins, L. M., Dziak, J. J., & Li, R. (2009). Design of experiments with multiple independent variables: A resource management perspective on complete and reduced factorial designs. *Psychological Methods*, *14*, 202–224. doi:10.1037/a0015826
- Coxon, A. P. M. (1999). *Sorting data: Collection and analysis*. Thousand Oaks, CA: Sage.
- Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan executive function system (D-KEFS)*. San Antonio, TX: Psychological Corp.
- DeSarbo, W. S. (1982). GENNCLUS: New models for general nonhierarchical clustering analysis. *Psychometrika*, *47*, 449–475.
- Gopnik, A., & Meltzoff, A. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development*, *58*, 1523–1531.
- Gopnik, A., & Meltzoff, A. N. (1992). Categorization and naming: Basic-level sorting in eighteen-month-olds and its relation to language. *Child Development*, *63*, 1091–1103.
- Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*, 404–411.
- Greenberg, E. A., & McIsaac, M. S. (1984). SIMATRIX: A program using the SAS system for multidimensional scaling analysis of sorting data. *Behavior Research Methods, Instruments, & Computers*, *16*, 407–408. doi:10.3758/BF03202471
- Helsen, K., & Green, P. E. (1991). A computational study of replicated clustering with an application to market segmentation. *Decision Sciences*, *22*, 1124–1141.
- Hunt, R. R., & Seta, C. E. (1984). Category size effects in recall: The roles of relational and individual item information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 454–464. doi:10.1037/0278-7393.10.3.454
- Isen, A. M., & Daubman, K. A. (1984). The influence of affect on categorization. *Journal of Personality and Social Psychology*, *47*, 1206–1217.
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*, 313–323.
- Jolicœur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*, 243–275. doi:10.1016/0010-0285(84)90009-4
- Jørgensen, C. (1995). *Image attributes: An investigation* (Doctoral dissertation, Syracuse University). Retrieved from http://surface.syr.edu/it_etd/39/
- Klatzky, R. L., Lederman, S. J., & Reed, C. (1987). There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision. *Journal of Experimental Psychology: General*, *116*, 356–369.

- Lickel, B., Hamilton, D. L., Wieczorkowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology, 78*, 223–246. doi:10.1037/0022-3514.78.2.223
- Mogilner, C., Rudnick, T., & Iyengar, S. S. (2008). The mere categorization effect: How the presence of categories increases choosers' perceptions of assortment variety and outcome satisfaction. *Journal of Consumer Research, 35*, 202–215.
- Murphy, G. L., & Smith, E. E. (1982). Basic-level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior, 21*, 1–20.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57. doi:10.1037/0096-3445.115.1.39
- Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*, 184–188. doi:10.1177/0963721414531598
- Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26*, 303–343.
- Quine, W. V. (1969). *Natural kinds*. Amsterdam, The Netherlands: Springer Netherlands.
- Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge, UK: Cambridge University Press.
- Rorissa, A., & Iyer, H. (2008). Theories of cognition and image categorization: What category labels reveal about basic level theory. *Journal of the American Society for Information Science and Technology, 59*, 1383–1392.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439. doi:10.1016/0010-0285(76)90013-X
- Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9*, 283–294.
- Rosenberg, S., & Park Kim, M. (1975). The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research, 10*, 489–502.
- Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology, 38*, 495–553.
- Russell, J. A., & Bullock, M. (1985). Multidimensional scaling of emotional facial expressions: Similarity from preschoolers to adults. *Journal of Personality and Social Psychology, 48*, 1290–1298.
- Shulman, C., Yirmiya, N., & Greenbaum, C. W. (1995). From categorization to classification: A comparison among individuals with autism, mental retardation, and normal development. *Journal of Abnormal Psychology, 104*, 601–609.
- Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition, 22*, 377–386. doi:10.3758/BF03200864
- Stringer, P. (1967). Cluster analysis of non-verbal judgments of facial expressions. *British Journal of Mathematical and Statistical Psychology, 20*, 71–79.
- Thomas, J. S., Ditzfeld, C. P., & Showers, C. J. (2013). Compartmentalization: A window on the defensive self. *Social and Personality Psychology Compass, 7*, 719–731. doi:10.1111/spc3.12061
- Towler, A. J., & Schneider, D. J. (2005). Distinctions among stigmatized groups. *Journal of Applied Social Psychology, 35*, 1–14.
- Tversky, A., & Gati, I. (1978). Studies of similarity. *Cognition and Categorization, 1*, 79–98.
- Ülkümen, G., Chakravarti, A., & Morwitz, V. G. (2010). Categories create mind-sets: The effect of exposure to broad versus narrow categorizations on subsequent, unrelated decisions. *Journal of Marketing Research, 47*, 659–671.
- Vohs, K. D., & Baumeister, R. F. (2004). Understanding self-regulation: An introduction. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 1–9). New York, NY: Guilford Press.