



Published in final edited form as:

J Exp Theor Artif Intell. 2009 September 1; 21(3): 197–215. doi:10.1080/09528130903065380.

The optimal level of fuzz: Case studies in a methodology for psychological research

Arthur B. Markman, Jennifer S. Beer, Lisa R. Grimm, Jonathan R. Rein, and W. Todd Maddox
University of Texas at Austin

Abstract

Cognitive Science research is hard to conduct, because researchers must take phenomena from the world and turn them into laboratory tasks for which a reasonable level of experimental control can be achieved. Consequently, research necessarily makes tradeoffs between internal validity (experimental control) and external validity (the degree to which a task represents behavior outside of the lab). Researchers are thus seeking the best possible tradeoff between these constraints, which we refer to as the *optimal level of fuzz*. We present two principles for finding the optimal level of fuzz, in research, and then illustrate these principles using research from motivation, individual differences, and cognitive neuroscience.

A hallmark of cognitive science involves the interplay of methods from different disciplines. Despite the importance of this interplay, methodological discussions in Psychology under the banner of cognitive science tend to focus on statistical issues such as the possibility that null hypothesis testing may lead research astray (e.g., Killeen, 2006). Much less discussion has centered on how to use the power of multidisciplinary cognitive science to construct research questions in ways that are likely to provide insight into the difficult questions that the field must address. In this paper, we present a principle that we call the *optimal level of fuzz* that we believe can guide good research. We start by defining the concept of fuzz and then discuss a set of principles that can guide researchers toward finding the optimal level of fuzz for their research. Next, we present three case studies of the optimal level of fuzz in action. Finally, we discuss the implications of this principle for research.

What is Fuzz?

Within cognitive science, experimental research in psychology provides data that can be used to constrain theories in neuroscience and psychology and to inspire new computational methods in artificial intelligence and reinforcement learning. Experimental research in psychology must typically trade off between internal and external validity. Internal validity is the basic idea that our experiments should be free from confounds and alternative explanations so that the results of our experiments can be unambiguously attributed to the variables we manipulated in our studies. External validity is the degree to which our studies reflect behavior that might actually occur outside the laboratory.

Logically speaking, of course, these dimensions do not have to trade off, but often human behavior is determined by many different interacting factors. Furthermore, people's mental states are not observable, and there are limits to what we can learn about the mind from observations that are restricted to the final results of a psychological process. Consequently, psychology has developed a variety of methods for manipulating circumstances in the

laboratory to bring about desired behaviors and has constructed systems for illuminating internal mental processes. For example, many studies use lexical decision tasks in which a subject is shown strings of letters that may or may not form a word and is asked to judge as quickly as possible whether the string forms a word. This task is quite useful for measuring the activity of concepts during a cognitive process. Tasks like this have been used in a variety of studies ranging from work on language comprehension to studies of goal activation (e.g., Fishbach, Friedman, & Kruglanski, 2003; McNamara, 2005). However, people are rarely asked to judge explicitly whether a string of letters is a word outside of the lab, and so this task does not reflect the kinds of behaviors that people carry out in their daily lives.

Optimizing Fuzz

Psychological research sits at the frontier of this tradeoff between internal and external validity, because it is addressing the thorniest questions about human behavior. After all, if a process could be understood just by observing it in the world, then there would be no reason to carry out research. Furthermore, there is a gradual accretion of knowledge in science so that previous research helps us understand psychological mechanisms that were once poorly understood.

However, researchers can choose a location on this frontier for carrying out research. Maximizing what is learned about a process requires both ensuring that studies address people's behavior outside the lab and also that studies strive to use tasks with high internal validity. We refer to this ideal as seeking the *optimal level of fuzz*. Although this may seem obvious, much of research is either too fuzzy or not fuzzy enough. At times, we might choose to observe a behavior that has not typically been studied, using very little experimental intervention, in order to better characterize the phenomena in the world that research aims to understand. Such research is often too fuzzy to adequately evaluate the processes underlying the real-world behavior. At other times, we might develop highly constrained laboratory studies that provide a high degree of control over the situation, thereby providing evidence for causal factors that influence performance. These studies are often not fuzzy enough to be sufficiently relevant to the original behavior of interest.

We believe that the multidisciplinary approach central to cognitive science provides a way to find the optimal level of fuzz. In this section, we discuss two principles that we believe are critical for finding the optimal level of fuzz:

1. The research problem must have some relationship to behavior outside the lab.
2. The research task must have a clear task analysis, preferably embedded in a mathematical, computational, or dynamical model.

We start with the principle that *research problems must bear some relationship to behavior outside the lab*. It is particularly important to try to minimize the number of layers of psychological research between the behavior in the lab and the behavior of interest outside the lab. For example, hundreds of psychological studies have been done on Tversky and Kahneman's (1983) "Linda" problem that was developed to study the conjunction fallacy.¹ This problem was an interesting demonstration of an apparent error in logical reasoning. It is clearly important to get to the root of factors that cause these apparent errors in judgment, but much of the research that has been done seems more focused on the Linda problem itself rather than on the broader psychological mechanisms governing reasoning behavior. In our view, at

¹The "Linda" problem is usually phrased like this: Linda is 28 years old. She is active in a number of women's rights groups. She volunteers in a shelter for battered women, and often participates in marches for abortion rights. Then, participants are asked to judge the probability that Linda is a bank teller and also the probability that she is a bank teller who is also a feminist. In the original studies, participants often judged the probability that she was a bank teller who is also a feminist to be higher than the probability that she is a bank teller even though the probability that she is a bank teller *cannot* be less than the probability that she is a bank teller who is also a feminist. This pattern of responses is called the *conjunction fallacy*.

the point where a line of research is focused more on a particular task from the literature than on a core phenomenon that occurs outside the lab, it is insufficiently fuzzy. That is, it does not have enough external validity to illuminate behavior outside the lab.

Part of the difficulty with implementing this principle is that laboratory research does have to explore phenomena to some level of depth to ensure that they are well understood. One good heuristic for determining whether research is sufficiently fuzzy is to focus on the degree to which the results of the study are going to be applied to the broader phenomenon of interest and to what degree they have implications only for the laboratory paradigm. For example, many studies have examined people's ability to learn to classify sets of simple items into one of a small number of categories (Shepard, Hovland, & Jenkins, 1961; Wisniewski, 2002). Studies within this paradigm have been used to explore whether people's category representations consist of prototypes or individual category exemplars (Medin & Schaffer, 1978; Posner & Keele, 1970). The conclusions of these papers were focused primarily on classification abilities outside the lab. However, there are also many studies that focus almost exclusively on adjudicating between particular mathematical models that apply only to this particular laboratory paradigm (e.g., Nosofsky & Palmeri, 1997; Smith, 2002). When the conclusion of a paper refers primarily to theories that explain behavior in the laboratory paradigm, then we take this as a signal that the research is insufficiently fuzzy.

The biggest gains in research, then, are going to come from examining important phenomena that are likely to occur outside the laboratory but are not well understood. Treading new ground has great potential for helping the field to better understand human behavior. To maximize the effectiveness of this research, however, it is crucial to follow the second key principle of optimal fuzz, which is *to embed research in laboratory tasks that are well understood*.

To make this principle clearer, we need to distinguish between tasks that are well-understood and those that are highly-studied. Highly-studied tasks have been employed in a large number of studies; well-understood tasks are characterized by the presence of a task analysis that specifies the cognitive processes brought to bear on the task and, often times, a set of models to characterize performance. For example, anagram-solving is a highly-studied task. However, there is no task analysis that describes how people go about solving anagrams, and so it is difficult to know what aspect of the process might be influenced by experimental manipulations. Thus, anagram solving is not an optimal task to use to study the effects of some other fuzzy variable (i.e., one that is related to behavior outside of the laboratory (see Principle 1)) on cognitive performance.

Good task analyses provide well-understood tasks with two important advantages for optimizing the level of fuzz. First, these task analyses support the development of models that can be used to describe people's performance. This is where techniques from cognitive science play a role. Although not obligatory, well-understood tasks often have mathematical, computational, or dynamical models that can be used to characterize the processes used by individuals through an analysis of their data. Models are quite useful, because they allow the researcher to go beyond gross measures of performance such as overall accuracy or response time that are often used as the primary results of psychological research. Often, there are many ways to achieve a particular level of performance, at the aggregate or even the trial by trial level. Models fit to the data from individual subjects make use of more information about their pattern of responses to characterize their performance (Maddox, 1999).

In addition, a good task analysis permits a statement about the optimal strategy in a particular task (Anderson, 1990; Geisler, 1989). In many studies, we give people a task and focus on the accuracy of their performance. We assume that higher accuracy is better than lower accuracy. In any given environment, however, there may be a particular strategy that is optimal to use.

Thus, we should be interested in whether subjects are using the optimal strategy for a particular environment (even if other cognitive limitations prevent them from achieving optimal accuracy using that strategy). As we will see, the ability to identify the optimal strategy in a given environment also allows us to distinguish between whether a particular psychological variable affects a person's propensity to use a particular type of strategy or whether it affects their ability to adapt to the needs of a particular environment.

This combination of fuzzy variables (Principle 1) with un-fuzzy tasks (Principle 2) is the paradigmatic case of the optimal level of fuzz. By embedding poorly understood variables within well-understood tasks, we minimize the amount of leeway in the interpretation of these studies. This strategy enables researchers to maximize the opportunity to turn fuzzy variables into unfuzzy variables. In the next section, we present three examples of research programs that have adopted a strategy that we see as consistent with the optimal level of fuzz.

The optimal level of fuzz in practice

We briefly review three lines of research that have used an optimal level of fuzz approach. The first is a systematic exploration of the role of incentives on cognitive performance. The second is an examination of individual differences in cognitive performance. The third comes from fMRI studies of cognitive processing. In each of these case studies, we discuss how the two principles of the optimal level of fuzz contributed to the success of the research effort. Two of these examples use tasks for which there are mathematical models that define optimal performance. The third uses computational modeling techniques from artificial intelligence to create a task model.

We make one comment at the outset, however. These sections focus on examples in which fuzzy variables become less fuzzy by using well-understood tasks. It is crucial to recognize that this strategy requires researchers to move out of their comfort zone in research. People who typically study highly controlled laboratory studies rarely venture into the realm of fuzzy variables. People who often study fuzzy variables rarely make use of laboratory tasks that can be characterized by solid task analyses or models. Thus, this style of research requires that some researchers make their work more fuzzy and others make their research less so. Indeed, to foreshadow a conclusion we reach at the end, this style of work rewards collaborations between individuals who have different styles of research.

Motivation and Incentives in Learning

Much research has been interested in the influence of incentives as motivating factors for cognitive performance (e.g., Brehm & Self, 1989; Locke & Latham, 2002). Much of this work has looked at organizational settings in which it becomes clear that incentives can influence performance, but it is quite difficult to determine the source of these effects. That is, work in these organizational contexts has high external validity, but it is hard to determine why these effects are obtained.

Other research has used incentives as a way of creating situational manipulations of motivational variables (Crowe & Higgins, 1997; Shah, Higgins, & Friedman, 1998). This work has allowed incentives to be manipulated as an independent variable, and so it aims at a higher degree of internal validity. However, most of the tasks used in this research are ones for which there are not good task analyses, and so it ultimately has a suboptimal level of fuzz. For example, Shah, Higgins, and Friedman (1998) had people solve anagrams under different incentive conditions. As discussed above, it is not clear what strategies people use to solve anagrams, and so only gross measures of performance can be obtained such as the number of anagrams solved correctly by participants. Thus, this work aimed for high internal validity, but the tasks were not understood well enough to be able to model the performance of individual

participants. Using our framework, this work meets the aim of Principle 1 but not Principle 2 resulting in a suboptimal level of fuzz.

In order to explore the influence of incentives on performance more systematically, we first turned to classification learning studies (see Maddox, Markman, & Baldwin, 2006). In these simple classification tasks, participants were shown lines that differed in their length, orientation, and position on a computer screen. They had to correctly learn to classify these items into one of two categories.

Classification tasks on their own often have a suboptimal level of fuzz, because they address issues that are primarily relevant to the study of laboratory classification tasks per se (i.e., they fail on Principle 1). However, they have a number of desirable properties that prove useful for generating an optimal level of fuzz when combined with poorly understood variables that have high external validity. First, there are mathematical models that can be used to assess the strategies that individual participants are using to make their classification decisions (F. G. Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Maddox, 1999). These models can be fit on particular blocks of trials during the study to assess how the strategy being used changes over the course of the study.

Second, different ways of structuring the categories being learned make different kinds of strategies optimal for the learning task. For example, if the rule that distinguishes between the two categories can be stated verbally (e.g., the long steep lines are in one category and the rest of the lines are in the other category), then the optimal strategy is to perform explicit hypothesis testing on different rules until the correct rule is found. Figure 1a shows a three-dimensional category structure. Each dimension of this “stimulus space” represents a perceptual dimension along which the items can vary. In this case, the items are lines that vary in their length, orientation, and position on a computer screen, and the points specify stimuli with particular lengths, orientations, and positions that were used in the study. Subjects given this category structure should use hypothesis testing to find this subtle verbal rule that combines length and orientation.

Not all rules can be stated verbally. For example, Figure 1b shows a category structure with a category boundary that is a line that cuts through the length and orientation dimensions. The best verbal description of this rule is “If the length of the lines is greater than the orientation, then the item is in one category, otherwise it is in the other category.” This statement does not really make sense, because length and orientation are measured in different units. Much research demonstrates that the best way to learn categories with this kind of *information integration* structure is to allow the implicit habit-learning system to slowly acquire the appropriate item-response mapping (e.g., Maddox & Ashby, 2004; Maddox, Ashby, & Bohil, 2003; Maddox, Filoteo, Hejl, & Ing, 2004). That is, people learning these categories must learn to respond using their ‘gut feeling’ about which response is correct. Thus, there are different optimal strategies for verbal rule-based and nonverbal information integration category structures.

Given this framework, it is possible to compare the influence of two different kinds of incentives, which we can loosely describe as *carrots* and *sticks*. A carrot is an incentive in which good performance is rewarded by some bonus. Higgins (1997) suggests that a carrot induces a motivational state that he calls a *promotion focus*, in which people become more sensitive to potential gains in their environment. A *stick* is an incentive in which something is taken away if performance is not good. Higgins (1997) suggests that sticks induce a motivational state called a *prevention focus*, in which people become more sensitive to potential losses in their environment.

In addition to manipulating incentives, we can manipulate the reward structure of the task. Most psychology experiments use a *gains* reward structure. Either participants receive points as rewards for performance or they are given positive feedback for correct responses. However, it is also possible to have a *losses* reward structure in which participants lose points during the study (though they presumably lose fewer points for correct responses than for incorrect responses).

Motivation from incentives is a fuzzy concept that can then be brought into an experimental session to be studied. Shah, Higgins, and Friedman (1998) gave people either a carrot (i.e., a situational promotion focus) or a stick (a situational prevention focus) and then asked them to solve anagrams. Anagrams printed in green would give them potential gains in the task and anagrams printed in red would give them potential losses in the task. Shah et al. found that participants given a carrot tended to solve gains anagrams and participants given a stick tended to solve loss anagrams.

There are many possible explanations for this effect. People might have their attention drawn to items whose reward matches the motivational state induced by the incentive. Perhaps participants are better able to adapt their strategy to rewards that fit the incentive. Finally, it is possible that a match between an incentive and the rewards of the task influence the kinds of cognitive processes that can be brought to bear on a task.

We explored these issues systematically using the classification methodology just described (Maddox, Baldwin, & Markman, 2006). Participants were offered either a carrot (an entry into a raffle to win \$50 if their performance exceeded a criterion) or a stick (they were given an entry into a raffle upon entering the lab and were told that they could keep the ticket if their performance exceeded a criterion). Some participants gained points during the task, and were given more points for a correct answer than for an incorrect answer. Some participants lost points, but they lost fewer points for a correct answer than for an incorrect answer. In one set of studies, participants learned a complex rule-based set of categories, while in a second set of studies, participants learned an information integration set of categories in which the rule distinguishing the categories could not be verbalized. In each study, mathematical models were fit to individual subjects' performance to examine how they changed strategies over the course of the study.

The pattern of data from these studies revealed that when the motivational state induced by the incentive matched the reward structure of the task, participants were more flexible in their behavior than when the motivational state mismatched the reward structure. That is, participants exhibited flexible behavior when the incentive was a carrot and the tasks gave them gains or when the incentive was a stick and the tasks gave them losses. Participants' behavior was relatively inflexible when the incentive mismatched the task. Flexibility was defined as the ability to explore potential rules in this task.

Of importance, whether a participant performed a particular task well depended on whether the task was one for which flexible behavior was advantageous. The rule-based task we used required flexibility, and so participants performed better when there was a fit between the incentive and the reward. Mathematical models fit to the data from this study suggested that participants with a motivational match between incentive and reward found the complex rule distinguishing the categories more quickly than did participants with a motivational mismatch.

In contrast, the information integration task required participants to avoid using rules and to allow the habit-learning system to learn the task. For this task, participants' performance was better when they had a mismatch between the incentive and the reward than when they had a match. In this case, the mathematical models suggested that participants with a motivational match persisted in using verbalizable rules in this task longer than did participants with a

mismatch, which hindered their ability to learn the categories (see Grimm, Markman, Maddox, & Baldwin, 2008, for a similar pattern of data).

This conclusion was corroborated by studies in a choice task. In the choice task, participants were shown two piles of cards (Worthy, Maddox, & Markman, 2007). Participants were told that the cards had point values on them. Their task was to draw cards from the decks. In the gains task, the cards had positive point values and participants had to maximize the number of points they obtained. In the losses task, the cards had negative point values, and participants had to minimize their losses. There are two general strategies that one can use in this task. An *exploratory* strategy involves selecting cards from each of the decks with a slight tendency to favor the deck that is currently yielding the best values. An *exploitative* strategy involves strongly favoring the deck that is currently yielding the best values. Pairs of decks were created that favored each type of strategy. That is, for some pairs of decks the participant had to be exploratory to exceed the performance criterion, while for other pairs of decks the participant had to be exploitative to exceed the performance criterion. It is possible to fit mathematical models to people's performance in this task as well (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006).

Consistent with the results of the classification task, we found that participants with a match between the motivational state induced by the incentive and the rewards in the task were more exploratory overall than were those who had a mismatch between incentive and reward. In particular, participants with a motivational match performed better on the decks that required exploration than on the decks that required exploitation. Participants with a mismatch performed better on decks that required exploitation than on decks that required exploration. Furthermore, the mathematical models fit to people's data have one parameter that measures a person's degree of exploration/exploitation. The value of this parameter was consistent with greater exploration for participants with a motivational match than for participants with a motivational mismatch for both types of tasks.

To summarize, then, there is a complex three-way interaction between a motivational state induced by an incentive, the reward structure of a particular task, and whether the task is best performed by flexible exploration of the environment. Previous research with a suboptimal level of fuzz had difficulty uncovering this pattern. By examining a poorly understood phenomenon (motivation and incentives) using very well-understood tasks, it was possible to discover this interaction. This work took tasks that are typically used in research that is not sufficiently fuzzy and brought it to bear on fuzzy variables in a manner that illuminated a complex interaction. To be clear, this work informs not only work on motivation and incentives but also work on classification and choice. Motivation and incentives change the strategies implemented by participants in these tasks. A complete computational model of classification should therefore include some component that corresponds to motivational state, the relevance of which would not have been apparent without placing classification research in the context of fuzzier motivational influences.

Individual differences in cognitive processing

There is growing interest in individual differences in cognitive performance. This interest arises from two sources. First, there have been intriguing observations of cultural differences in cognitive performance that suggest the results of studies done on Western college students do not generalize to all populations (e.g., Nisbett, Peng, Choi, & Norenzayan, 2001; Peng & Nisbett, 1999). On the basis of these observations, researchers have explored individual difference variables that are correlated with cultural differences that might explain these differences in performance (Gardner, Gabriel, & Lee, 1999; Kim & Markman, 2006).

For example, Nisbett and his colleagues have conducted a fascinating set of studies documenting cultural differences in cognition (Masuda & Nisbett, 2001; Peng & Nisbett, 1999). They find that East Asian participants have a greater relative preference for proverbs that embody contradictions than do Westerners. East Asians tend to give solutions to dilemmas that involve compromise more than do Westerners. Recognition memory for objects in East Asians is more influenced by the background context in which the object appears than are Westerners. Nisbett has argued that these differences in performance may reflect a long history of differences in cultural evolution between East Asian and Western cultures (Nisbett, Peng, Choi, & Norenzayan, 2001).

Second, studies have demonstrated that individual differences correlated with cultural differences affect cognitive performance. One such difference is self-construal (Markus & Kitayama, 1991). Much research has demonstrated that East Asians tend to represent themselves by focusing on the relationship between self and others. For example, the self-concept may be populated with the roles the self plays in relation to others, like teacher and friend. That is, East Asians have an *interdependent* self-construal. In contrast, Westerners tend to represent themselves by focusing on characteristics of themselves as individuals. In this case, the self-concept contains individual attributes, such as brown hair and likes golf. That is, Westerners tend to have an *independent* self-construal.

Despite these chronic differences in self-construal across cultures, it is possible to temporarily prime an interdependent or an independent self-construal. For example, Gardner and colleagues (Gardner, Gabriel, & Lee, 1999; Lee, Aaker, & Gardner, 2000) have used a technique in which they have participants circle the pronouns in a passage. An independent self-construal can be primed by having participants circle the pronouns “I” and “me.” An interdependent self-construal can be primed by having participants circle the pronouns “we” and “our.”

Kuhnen and Oyserman (2002) examined the influence of self-construal on cognitive performance. They found systematic shifts in people’s performance with a manipulation of self-construal. For example, they showed people large letters made up of smaller letters like the example in Figure 2 that shows the letter F made up of smaller letter H’s (Navon, 1977). They asked people either to identify the small letters or the large ones. Participants primed with an independent self-construal identified the small letters faster than the large letters, while participants primed with an interdependent self-construal were faster to identify the large letters than the small letters. They suggest that results like this are consistent with the possibility that individuals with an interdependent self-construal are more sensitive to context than are those with an independent self-construal.

A difficulty interpreting studies like this (as well as the studies by Nisbett and colleagues) is that there is no generally accepted task analysis for most of the studies done on cultural differences and self-construal. Thus, there are many reasons why someone might identify the smaller letters more quickly than the larger ones (or vice versa), to prefer proverbs that embody contradictions to those with no contradiction, or to be relatively more strongly influenced by background context. For example, speed of identification of small and large letters could be due to context sensitivity (as proposed by Kuhnen and Oyserman (2002)), or it could be a result of shifts in attention from global visual form to fine detail information that reflect information processing channels within the visual system (Oliva & Schyns, 1997; Sanocki, 2001). Thus, there is still a high degree of fuzz within this research area (because it meets Principle 1 but not Principle 2). To examine this issue more carefully, studies were done exploring the influence of differences in self-construal on causal induction.

In causal induction tasks, people are shown a set of potential causes that could influence an effect (P. W. Cheng, 1997; Waldmann, 2000). For example, the potential causes might be

liquids that influence the growth of plants, and the effect could be the growth of flowers on the plant (Spellman, 1996). Participants typically see many experimental trials that pair potential causes with effects. While induction from observed events can occur outside of the laboratory, individuals do not often encounter and assess causality from dozens of pairings of causes and effects. As such, this research area tends to be insufficiently fuzzy (because it meets Principle 2 but not Principle 1). An advantage of using this domain to study individual differences is that there are mathematical models that can be used to determine optimal behavior in this domain, and models that can help to assess what information participants are using when they perform this task (e.g., P. W. Cheng, 1997; Novick & Cheng, 2004; Perales, Catena, Shanks, & Gonzalez, 2005).

Typically, people use some kind of covariation information between causes and effects to assess the influence of a potential cause on the effect. If an effect occurs more often in the presence of a cause than in its absence, that cause is likely to promote the effect. If an effect occurs less often in the presence of a cause than in its absence, that cause is likely to inhibit the effect.²

A particularly interesting covariation structure is one that embodies *Simpson's paradox* (see Spellman, 1996). In Simpson's paradox, the presence of two causes is confounded so that both of them co-occur often. The effect also occurs often in the presence of both causes, and so people often infer that both causes have the potential to promote the effect. However, one of the causes actually inhibits the effect. The only way to recognize correctly that one of the causes is inhibitory is to pay attention to the rare cases in which the inhibitory cause occurs in the absence of the other cause (which promotes the effect). That is, correctly identifying which cause promotes the effect and which cause inhibits the effect requires paying attention to the context in which the causes appear. Thus, the optimal strategy in this task is to pay attention to the context in which a potential cause appears. Further, it is possible to mathematically calculate the relationship between a cause and an effect taking the other cause into account, which is known as a conditional contingency (Spellman, 1996). Using mathematical models, one can determine if a given subject's causal judgment or predictions during learning approximate an unconditional contingency (e.g., attention to a single cause) or a conditional contingency (e.g., attention to the context).

Kim, Grimm, & Markman (2007) exposed people to a covariation structure embodying Simpson's paradox following a self-construal prime. A third of the participants were primed to have an interdependent self-construal, a third were primed to have an independent self-construal, and the remaining third did not get a prime and served as a control group. Consistent with the idea that people with an interdependent self-construal are more influenced by context than are those with an independent self-construal, people primed with an interdependent self-construal were more likely to recognize that the inhibitory cue was indeed an inhibitor of the effect than were those with an independent self-construal or the control participants.

This example demonstrates further how using the optimal level of fuzz can help us understand the influence of a poorly understood variable on cognitive performance using a task that is often examined only in very un-fuzzy contexts. In this case, individual differences appear to influence behavior outside of the laboratory (Principle 1). However, there were many

²Mathematically, this relationship between a cause and an effect is formalized in the ΔP rule. In this rule, the unconditional contingency (UC) between a cause (C) and an effect (E) represents the covariation between the cause and the effect ignoring other potential causes. This UC is computed by subtracting the probability of the effect when the cause is absent from the probability of the effect when the cause is present:

$$\Delta P = P(E|C) - P(E|\sim C)$$

speculations about the effects of individual difference variables like self-construal on performance. By using a causal reasoning task (which often appears in studies whose implications are drawn primarily for other laboratory causal reasoning tasks), it is possible to generate a rigorous mathematical definition of sensitivity to context (Principle 2). Therefore, researchers can demonstrate more conclusively that manipulations of self-construal affect the degree to which people are influenced by contextual variables.

The optimal level of fuzz in cognitive neuroscience

The final example we present in this paper examines the interpretation of fMRI data. Understanding how the brain implements any number of psychological processes could not be more fundamental in terms of focusing research on phenomena that have real-world relevance. Neuroscience methodologies such as fMRI have been used for some time to understand perceptual and cognitive processing; more recently these methodologies have been used to investigate how the brain implements social and affective processes as well.

Studies using fMRI most commonly examine differences in brain activity by comparing experimental conditions that show significant behavioral differences. Many studies do not statistically examine the relation between brain activations and behavioral differences. Therefore, many fMRI experiments rely on gross measures of behavior and they may tell us more about where a process might occur in the brain instead of how that brain region implements that process (e.g., Buxton, 2002; Poldrack & Wagner, 2004).

A number of promising lines of research, however, have begun to use techniques that follow the principles of the optimal level of fuzz. In this research, people are given tasks for which there is a good task analysis. This task analysis is embedded in a mathematical or computational model, and that model is then fit to people's behavior. The parameters that emerge from this modeling effort are then related to changes in neural activity, as measured by the relative amount of oxygenated blood (BOLD signal), in particular brain regions to better understand how they implement psychological processes. In this section, we give three brief examples of this technique in action. The first predicts activity in multiple brain regions from different values of a single model parameter. The second uses a model to separate apparently similar behavior into qualitatively different processes, corresponding with differing activity levels in identical regions. The last splits activity according to the timing of different model computations, predicting the complex time-course of several brain areas within a single trial.

Gläscher and Büchel (2005) explored how different brain regions underlie learning in Pavlovian conditioning. In particular, they were interested in whether different areas support learning at different time-scales. Previous neuroimaging studies could not disambiguate short- and long-term learning because they did not vary the conditioned stimulus-unconditioned stimulus (CS-US) contingency (e.g., D. T. Cheng, Knight, Smith, Stein, & Helmstetter, 2003; Knight, Cheng, Smith, Stein, & Helmstetter, 2004). Gläscher and Büchel avoided this ambiguity by creating a situation in which CS-US contingencies varied over time. Two pictures of a face and house oscillated between being strongly and weakly predictive of a painful stimulus. People reliably learned when the painful stimulus would follow a given picture, as measured by skin conductance and explicit predictions.

Of course, how participants predicted the US is the fuzzy issue. Any single correct prediction—and the corresponding BOLD signal—is due to the interaction of short- and long-term learning. To tease the two apart, Gläscher and Büchel used a mathematical model with a variable learning rate parameter. In the Rescorla-Wagner model (Rescorla & Wagner, 1972), learning on a single trial is a function of previous predictions and the prediction error on that trial. The weight placed on the prediction error influences how rapidly the model responds to a single trial; low and high weights represent slow and fast learning rates, respectively. Gläscher

and Büchel compared model predictions with slow and fast learning rates separately against fMRI data from different brain regions. They found that slow-learning model predictions matched activity in the amygdala, an area associated with maintaining stimulus-response contingencies. In contrast, they found that fast-learning model predictions matched activity in fusiform and parahippocampal cortices, areas associated with visual processing and more short-term memory phenomena, such as priming. Without the differing model predictions, they would have been unable to dissociate the behavioral relevance of the activity in these regions. Worse yet, without the finely-tuned functions to track signal within the fuzzy noise, they may not have found significant activity in either region.

Another case of using modeling to simplify complex activity in a complex environment comes from work on the different processes underlying decision making under uncertainty (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006). Daw et al. scanned participants while they chose between four potential rewards. Reward on each trial was probabilistic, and the relative probabilities of the four options changed over time. To optimize reward, one must often exploit the option that appears to have the highest probability of reward, but also explore other options that may have become more rewarding. Decision making in such a complex, variable environment is similar to naturalistic tasks like foraging. The variability also makes it difficult to separate exploitative decisions from exploratory ones.

Daw et al. used a reinforcement learning model, similar in principle to the Rescorla-Wagner model, to fit participants' data. At a given trial, the model uses the participant's previous decisions and reward history to represent a belief about the expected value of each option, the certainty of that belief, and the likelihood for exploration. Using these model estimates, decisions on individual trials can be separated according to whether a person selected the option with the highest expected value or another option. This separation allows for the isolation of the neural patterns underlying the exploration and exploitation patterns uniquely.

Daw et al. found that exploratory decisions (i.e. choices of sub-maximal options) are associated with increases in the BOLD signal in frontopolar cortex and intraparietal sulcus, areas associated with cognitive control and decision making. In contrast, exploitative decisions (i.e. choices of maximal options) are associated with decreases in signal in the same regions. Again, use of a well-understood task was critical for finding these differences. If the authors had treated each decision uniformly instead of using model-driven estimates of process, they would not have seen any differences.

In addition to analyzing tasks that involve different processes on different trials, it is also possible to use models to decompose the processes and neural signal within a single trial (Anderson & Qin, 2008). While most fMRI studies usually focus on simple tasks with trials that take place in a few seconds, Anderson and his colleagues explore more naturalistic reasoning tasks that may take minutes or hours (thereby meeting Principle 1). To look at such large-scale behavior, Anderson and Qin scanned people while they performed a task that involved completing a series of arithmetic operations on each trial. The task was designed so that all participants would go about performing the task using the same set of steps. This approach was used to eliminate as much variability across people as possible. They modeled participants' performance using principles of symbolic AI embedded in ACT-R (Anderson, 1993), a production system that acts like a programming language for constructing models of complex tasks. The model included six modules, each designed to perform a function associated with a different brain region, such as obtaining visual information about the problem from a computer screen, retrieving information from memory, and preparing motor operations to make responses (thus meeting Principle 2).

The model makes predictions about when and how long each module is active for each arithmetic problem. For example, in a problem that requires five intermediate responses, the motor module would be briefly active during those five moments. Anderson and Qin compared the model estimates of each module's activity against the BOLD signal in the brain regions associated with that module's type of processing. For example, they predicted that the left fusiform area would be associated with the type of visual processing relevant to the task. In the fMRI scans from their study, they found a correlation of .91 between activity in this brain area and the predictions of the model. High correlations were also obtained between the predictions of the model and activity in areas that they associated with memory retrieval and goal orientation.

Again, the model analysis is crucial for understanding the fMRI data. Without separating portions of the BOLD signal, activity in regions only briefly but critically involved in the task could have been undetectable. Likewise, regions that were active would have been lumped together in an undifferentiated network. Labeling these regions simply as "involved in arithmetic reasoning" belies their unique functions and the complexity of their interaction.

The approach embodied in these three studies represents an important advance in the use of fMRI to study behavior. The use of a detailed model of behavior that specifies a step-by-step process by which people carry out a task supports detailed predictions about the way that brain activity in different regions of the brain will vary over time. Such detail increases the likelihood of understanding the relation between neural activity and how it supports a particular psychological process. This approach allows researchers to make stronger claims about the functions of brain areas than are possible just by looking at brain regions that are active throughout the task globally. Thus, the optimal level of fuzz approach applied to cognitive neuroscience holds significant promise for allowing us to better understand the interactions among brain regions in cognitive processing.

Implications

While all researchers are likely to agree that finding an optimal level of fuzz is important in principle, it is difficult to achieve in practice. Thus, we identified two principles that can help lead to an optimal level of fuzz in research, and then we provided examples of these principles in action. First, research should seek to explain phenomena that are not yet well understood, but are focused on behaviors that occur outside the laboratory. Second, the experimental approach to these problems should seek tasks that are quite well understood and, ideally, permit models that can be fit to the data of individual participants. This latter point encourages researchers studying behavior to work with people in other areas of cognitive science to develop formal and computational models of task performance that can be used to illuminate patterns of data.

One important point that emerges from this work is that the optimal level of fuzz is best achieved by research teams rather than by individual researchers. It is unlikely that any individual researcher will have expertise in the variety of areas that are required to do this work well. For example, research on motivation requires familiarity with a range of literatures that do not often overlap with work on many cognitive tasks. Furthermore, not every researcher is likely to have the modeling skills necessary to fit the data from individual subjects to a model that is diagnostic of performance. Thus, this approach to research prizes collaborative work. The experience of the authors of this paper is that such collaborative work, while difficult to get started, has great rewards. Furthermore, collaborative work has always been at the heart of cognitive science as a discipline.

In addition, this style of work requires that researchers and reviewers be open to research areas and research techniques that cross traditional disciplinary lines. Again, this issue is one central to the core of cognitive science. The first two examples we presented in this paper are projects that combine topics frequently explored by social psychologists with research techniques drawn from the cognitive literature. The last example draws on expertise in brain imaging and cognitive psychology. Often, research like this pulls both the researchers and the readers of papers on these techniques out of their comfort zone. Indeed, researchers often have a visceral negative reaction to research that takes a strikingly different approach to work than the one they typically adopt. The optimal level of fuzz approach requires checking this gut reaction and embracing new styles of research.

The three examples given in this paper are all cases of ongoing research in which new insights are being generated by adopting new approaches to research. In addition, there are many examples of how introducing new methods has pulled researchers out of their comfort zone in ways that have led to important new results. For example, the introduction of data from neuroscience to the study of classification has had a beneficial impact on the field. By the early 1990s, much research in classification focused on comparisons of model fits to simple classification tasks, and many papers were focused on determining whether one model accounted for more data than another (e.g., F.G. Ashby & Maddox, 1993; Nosofsky & Palmeri, 1997; Smith & Minda, 2000). Researchers began to draw insights from neuroscience. This focus led to a recognition that there are multiple memory systems that may serve our ability to learn new categories, and that different types of categories may thus be learned in qualitatively different ways (e.g., F. G. Ashby, Alfonso-Reese, Turken, & Waldron, 1998). This research has led to important new discoveries using methods from cognitive neuroscience (such as studies of patient groups and imaging studies) as well as more traditional experimental methods (e.g., F.G. Ashby, Ennis, & Spiering, 2007; Knowlton, Squire, & Gluck, 1994; Maddox & Filoteo, 2001).

The optimal level of fuzz suggests that model development and task analysis is an important part of laboratory research studies. Often, psychologists bring a task into the lab, but then analyze the data only with ANOVA or simple regression models. These statistical techniques are models of performance that refer primarily to the experimentally manipulated variables in a study. Often, however, there is other information in people's performance that can be used to distinguish among alternative explanations of how people were carrying out a task. It is frequently worth the effort to develop some kind of model to help distinguish among these alternatives. Often, when it proves difficult to develop a model for a particular task, that is a sign that there are aspects of the task that are not yet well-enough understood to support model development. Rather than shying away from developing these task analyses, it is important to dive into this process. As we have seen, exploring fuzzy variables often informs model development by illuminating previously unforeseen complexities.

A further implication of the optimal level of fuzz is that researchers must question the assumptions underlying their typical research paradigms. Often, in an effort to control a paradigm, researchers start with a desired outcome, and then adjust the paradigm until the experiment "works." When the study works out far differently than expected, the researcher concludes that there was something wrong with the study and repeats it, tweaking the methodology to achieve the desired result.

Finally, it is important to recognize that the optimal level of fuzz is a process for research. Initially, every task worthy of study is poorly understood. Much research needs to be done to generate task analyses and to create models. Every task is fuzzy at first, and requires research to be made less fuzzy. At the point where a laboratory task is sufficiently well-understood, it takes substantial ingenuity to identify other fuzzy variables that are good candidates to be

brought into studies using well-understood tasks. Over time, of course, even these fuzzy variables become less fuzzy, thereby raising new opportunities for research to progress.

We suggest that there is much to be learned about human behavior from the parameters within which experimental procedures work as expected. Rather than adjusting our experiments just to ensure that they work as expected, we should use these parameters to inform our work as we seek to understand more complex elements of human behavior. For example, we often add motivational incentives to our research projects under the assumption that these incentives will get participants to “pay more attention” to our study or to “try harder.” We find a combination of incentives that gets participants to perform the task of interest to us. However, by treating these incentives and motivational processes as background parts of our experimental procedure, we fail to get a firmer grasp on important concepts like what motivational aspects of a task get people to “pay more attention” or what it means psychologically to “try harder.” On the optimal level of fuzz approach, these questions should be front and center in our research.

Acknowledgments

This research was supported by AFOSR grant FA9550-06-1-0204 and NIMH grant R01 MH0778 given to ABM and WTM. This work was also supported by a fellowship in the IC² institute to ABM. The authors thank the motivation lab group for their comments on this work.

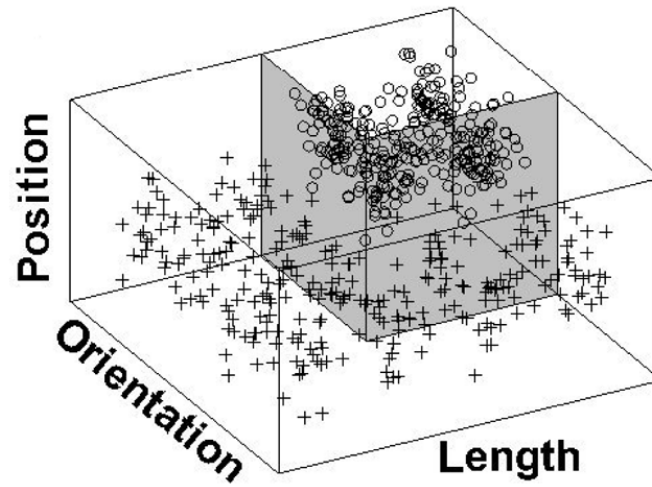
References

- Anderson, JR. *The Adaptive Character of Thought*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1990.
- Anderson JR, Qin Y. Using brain imaging to extract the structure of complex events at the rational time band. *Journal of Cognitive Neuroscience* 2008;20(9):1624–1636. [PubMed: 18345979]
- Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM. A neuropsychological theory of multiple systems in category learning. *Psychological Review* 1998;105(3):442–481. [PubMed: 9697427]
- Ashby FG, Ennis JM, Spiering BJ. A neurobiological theory of automaticity in perceptual categorization. *Psychological Review* 2007;114(3):632–656. [PubMed: 17638499]
- Ashby FG, Maddox WT. Relations between prototype, exemplar and decision bound models of categorization. *Journal of Mathematical Psychology* 1993;37:372–400.
- Brehm JW, Self EA. The intensity of motivation. *Annual Review of Psychology* 1989;40:109–131.
- Buxton, RB. *Introduction to functional magnetic resonance imaging: Principles and techniques*. New York: Cambridge University Press; 2002.
- Cheng DT, Knight DC, Smith CN, Stein EA, Helmstetter FJ. Functional MRI of human amygdala activity during Pavlovian fear conditioning: Stimulus processing versus response expression. *Behavioral Neuroscience* 2003;117:3–10. [PubMed: 12619902]
- Cheng PW. From covariation to causation: A causal power theory. *Psychological Review* 1997;104(2):367–405.
- Crowe E, Higgins ET. Regulatory focus and strategic inclinations: Promotion and prevention in decision-making. *Organizational Behavior and Human Decision Processes* 1997;69(2):117–132.
- Daw ND, O’Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. *Nature* 2006;441:876–879. [PubMed: 16778890]
- Fishbach A, Friedman RS, Kruglanski AW. Leading us not into temptation: Momentary allurements elicit overriding goal activation. *Journal of Personality and Social Psychology* 2003;84(2):296–309. [PubMed: 12585805]
- Gardner WL, Gabriel S, Lee AY. “I” value freedom, but “we” value relationships: Self-construal priming mirrors cultural differences in judgment. *Psychological Science* 1999;10(4):321–326.
- Geisler WS. Sequential ideal-observer analysis of visual discriminations. *Psychological Review* 1989;96:267–314. [PubMed: 2652171]
- Glaescher J, Buechel C. Formal learning theory dissociates brain regions with different temporal integration. *Neuron* 2005;47:295–306. [PubMed: 16039570]

- Grimm LR, Markman AB, Maddox WT, Baldwin GC. Differential effects of regulatory fit on category learning. *Journal of Experimental Social Psychology* 2008;44:920–927. [PubMed: 19412324]
- Higgins ET. Beyond pleasure and pain. *American Psychologist* 1997;52(12):1280–1300. [PubMed: 9414606]
- Killeen PR. Beyond statistical inference: A decision theory for science. *Psychological Bulletin* 2006;13(4):549–562.
- Kim K, Grimm LR, Markman AB. Self-construal and the processing of covariation information in causal reasoning. *Memory and Cognition* 2007;35(6):1337–1343.
- Kim K, Markman AB. Differences in fear of isolation as an explanation of cultural differences: Evidence from memory and reasoning. *Journal of Experimental Social Psychology* 2006;42:350–364.
- Knight DC, Cheng DT, Smith CN, Stein EA, Helmstetter FJ. Neural substrates mediating human delay and trace fear conditioning. *Journal of Neuroscience* 2004;24:218–228. [PubMed: 14715954]
- Knowlton BJ, Squire LR, Gluck MA. Probabilistic classification learning in amnesia. *Learning and Memory* 1994;1:106–120. [PubMed: 10467589]
- Kuhnen U, Oyserman D. Thinking about the self influences thinking in general: Cognitive consequences of salient self-concept. *Journal of Experimental Social Psychology* 2002;38:492–499.
- Lee AY, Aaker JL, Gardner W. The pleasures and pains of distinct self-construals: The role of interdependence in regulatory focus. *Journal of Personality and Social Psychology* 2000;78(6):1122–1134. [PubMed: 10870913]
- Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist* 2002;57(9):705–717. [PubMed: 12237980]
- Maddox WT. On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. *Perception and Psychophysics* 1999;61(2):354–374. [PubMed: 10089766]
- Maddox WT, Ashby FG. Dissociating explicit and procedure-learning based systems of perceptual category learning. *Behavioral Processes* 2004;66(3):309–332.
- Maddox WT, Ashby FG, Bohil CJ. Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2003;29(4):650–662.
- Maddox WT, Baldwin GC, Markman AB. Regulatory focus effects on cognitive flexibility in rule-based classification learning. *Memory and Cognition* 2006;34(7):1377–1397.
- Maddox WT, Filoteo JV. Striatal contributions to category learning: Quantitative modeling of simple linear and complex nonlinear rule learning in patients with Parkinson's Disease. *Journal of the International Neuropsychological Society* 2001;7(6):710–727. [PubMed: 11575593]
- Maddox WT, Filoteo JV, Hejl KD, Ing AD. Category number impacts rule-based but not information-integration category learning: Further evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2004;30(1):227–245.
- Maddox WT, Markman AB, Baldwin GC. Using classification to understand the motivation-learning interface. *Psychology of Learning and Motivation* 2006;47:213–250.
- Masuda T, Nisbett RE. Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology* 2001;81(5):922–934. [PubMed: 11708567]
- McNamara, TP. *Semantic priming: Perspectives from memory and word recognition*. New York: Psychology Press; 2005.
- Medin DL, Schaffer MM. Context theory of classification. *Psychological Review* 1978;85(3):207–238.
- Navon D. The precedence of global features in visual perception. *Cognitive Psychology* 1977;9:353–383.
- Nisbett RE, Peng K, Choi I, Norenzayan A. Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review* 2001;108(2):291–310. [PubMed: 11381831]
- Nosofsky RM, Palmeri TJ. Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification. *Perception and Psychophysics* 1997;59(7):1027–1048. [PubMed: 9360476]
- Novick LR, Cheng PW. Assessing interactive causal influence. *Psychological Review* 2004;111(2):455–485. [PubMed: 15065918]

- Oliva A, Schyns PG. Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology* 1997;34:72–107. [PubMed: 9325010]
- Peng KP, Nisbett RE. Culture, dialectics, and reasoning about contradiction. *American Psychologist* 1999;54(9):741–754.
- Perales JC, Catena A, Shanks DR, Gonzalez JA. Dissociation between judgments and outcome expectancy measures in covariation learning: A signal detection theory approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2005;31(5):1105–1120.
- Poldrack RA, Wagner AD. What can neuroimaging tell us about the mind? *Current Directions in Psychological Science* 2004;13(5):177–181.
- Posner MI, Keele SW. Retention of abstract ideas. *Journal of Experimental Psychology* 1970;83:304–308.
- Sanocki T. Interaction of scale and time during object identification. *Journal of Experimental Psychology: Human Perception and Performance* 2001;27(2):290–302. [PubMed: 11318048]
- Shah J, Higgins ET, Friedman RS. Performance incentives and means: How regulatory focus influences goal attainment. *Journal of Personality and Social Psychology* 1998;74(2):285–293. [PubMed: 9491583]
- Shepard RN, Hovland CL, Jenkins HM. Learning and memorization of classifications. *Psychological Monographs* 1961:75.
- Smith JD. Exemplar theory's predicted typicality gradient can be tested and disconfirmed. *Psychological Science* 2002;13(5):437–442. [PubMed: 12219810]
- Smith JD, Minda JP. Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2000;26(1):3–27.
- Spellman BA. Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science* 1996;7(6):337–342.
- Tversky A, Kahneman D. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 1983;90(4):293–315.
- Waldmann MR. Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2000;26(1):53–76.
- Wisniewski, EJ. Concepts and categories. In: Pashler, H.; Medin, DL., editors. *Stevens Handbook of Experimental Psychology*. Vol. 3. Vol. 2. New York: John Wiley and Sons; 2002. p. 467-531.
- Worthy DA, Maddox WT, Markman AB. Regulatory fit effects in a choice task. *Psychonomic Bulletin and Review* 2007;14(6):1125–1132. [PubMed: 18229485]

(a)



(b)

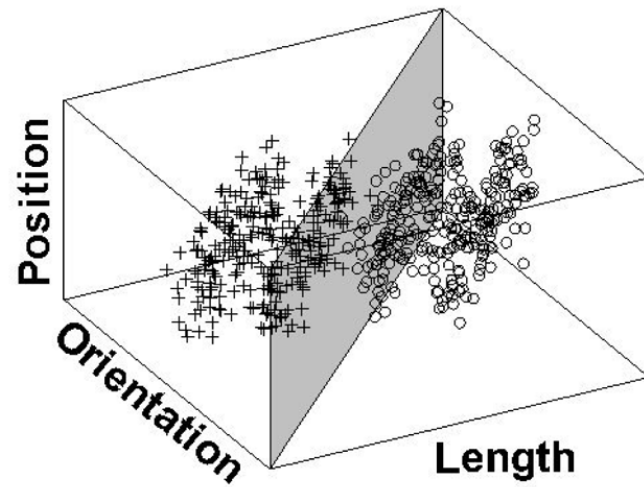


Figure 1.

(a) Sample rule-based category structure, and (b) sample information integration category structure. Rule-based category structures are easy to describe verbally. Information integration structures are hard to describe verbally.

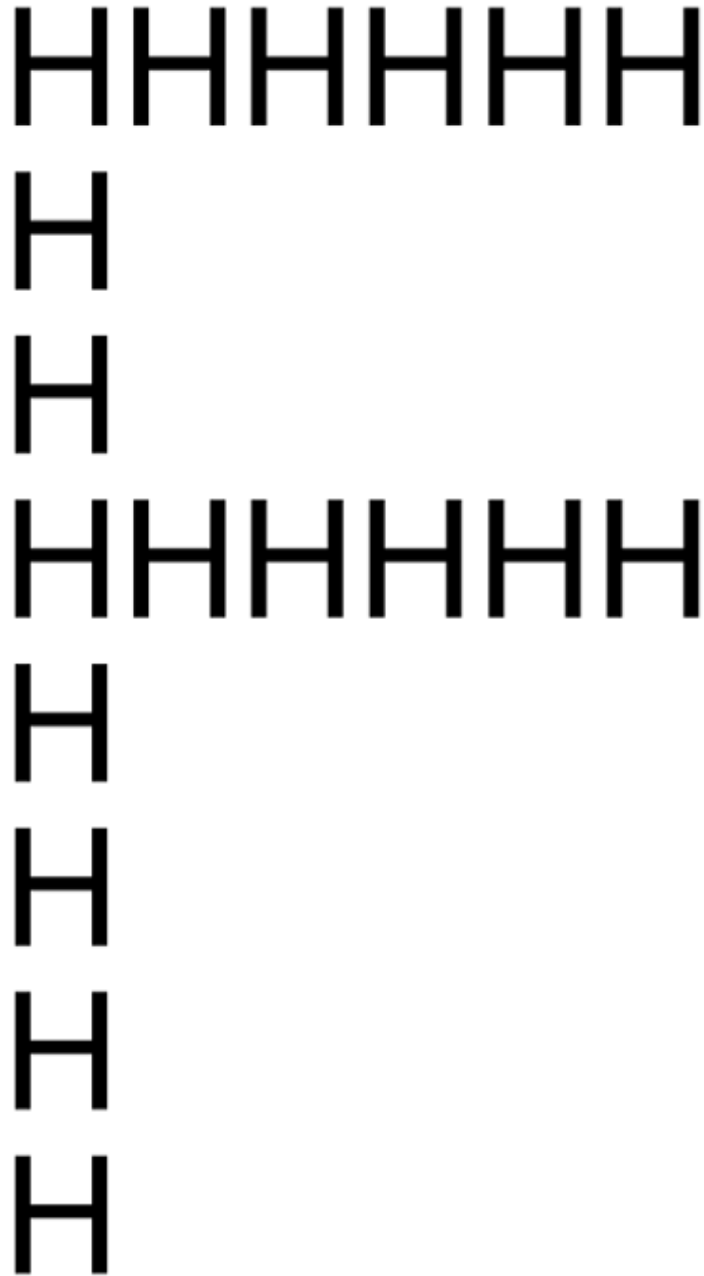


Figure 2. Sample stimulus with a large letter made up of smaller letters like those used in studies by Navon (1977). In this case, the large letter F is made of small letter H's.