An Examination of Learning Processes During Critical Incident Training: Implications for the Development of Adaptable Trainees

Andrew Neal
The University of Queensland

Terry Kirkpatrick and Graham Dewsnap
New South Wales Fire Brigades

Stuart T. Godley
University of Sydney

Wendy Joung and Beryl Hesketh
University of Sydney

Three experiments are reported that examined the process by which trainees learn decision-making skills during a critical incident training program. Formal theories of category learning were used to identify two processes that may be responsible for the acquisition of decision-making skills: rule learning and exemplar learning. Experiments 1 and 2 used the process dissociation procedure (L. L. Jacoby, 1998) to evaluate the contribution of these processes to performance. The results suggest that trainees used a mixture of rule and exemplar learning. Furthermore, these learning processes were influenced by different aspects of training structure and design. The goal of Experiment 3 was to develop training techniques that enable trainees to use a rule adaptively. Trainees were tested on cases that represented exceptions to the rule. Unexpectedly, the results suggest that providing general instruction regarding the kinds of conditions in which a decision rule does not apply caused them to fixate on the specific conditions mentioned and impaired their ability to identify other conditions in which the rule might not apply. The theoretical, methodological, and practical implications of the results are discussed.

Keywords: training, decision making, examples, rules, learning, fire fighting

Just over a decade ago, Ford and Kraiger (1995) argued for the need to apply cognitive constructs and principles to the analysis, design, and evaluation of personnel training programs. With the increasing cognitive complexity of work, Ford and Kraiger argued that there is a need to train for higher order cognitive skills. To do this, they argued, further research is needed to improve our understanding of the processes underlying learning. Since 1995, 41 articles have been published in the *Journal of Applied Psychology* with training, skill learning, or learning strategies as a keyword. Nine of these articles have focused specifically on cognitive constructs or principles. Of these, five have examined the role of self-regulatory processes in learning and/or transfer (e.g., Debowski, Wood, & Bandura, 2001; Holladay & Quinones, 2003), and four have examined the role of knowledge structures (e.g., Davis & Yi, 2004; Day, Arthur, & Gettman, 2001). None has examined the processes underlying learning.

In the current article we aim to address this gap at a theoretical, methodological, and practical level. Theoretically, we develop a model of learning processes in the context of a critical incident training program. In doing so, we draw on recent advances relating to the use of rules and examples in category learning (Markman & Ross, 2003; Rouder & Ratcliff, 2004) and apply them to the question of how trainees learn tactical decision-making skills. Methodologically, we aim to introduce a new technique to the personnel training literature that is based on this model of learning processes: process dissociation (Jacoby, 1998; Neal, Hesketh, & Andrews, 1995). Practically, we aim to use the results of the process dissociation procedure as a basis for designing training programs that reduce the incidence of trainees inappropriately generalizing rules and examples to cases in which they do not apply. This article, therefore, brings new theory and methodology to bear on the analysis, design, and evaluation of training programs.

The theoretical aims of this article derive from the growing interest in the way that people make decisions in complex, uncertain, dynamic environments (Endsley, 1995; Klein, 1993). Studies of naturalistic decision making suggest that people typically do not evaluate explicitly a set of decision alternatives, as assumed in expected utility models. Instead, this research suggests that individuals assess the situation relatively rapidly, recognize the type of problem involved, and retrieve a potential solution from memory (Lipshitz, Klein, Orasanu, & Salas, 2001). This style of decision making is known as recognition-primed decision making (Klein, 1993). This naturalistic decision-making literature has provided important qualitative insights into the way that experts make decisions in complex, uncertain, dynamic environments. However,
very little is known about the way in which trainees learn these decision skills.

Our theoretical contribution is to integrate the qualitative insights obtained from the naturalistic decision-making literature with a formal theoretical account of the learning process found in the categorization literature. The reason for drawing on theories of category learning is that recognition-primed decisions are inherently categorical ("Is this a problem of type x?"). Indeed, a number of authors have advocated using techniques that are very similar to those used in experimental studies of category learning when training in decision skills (Cohen et al., 2000; Klein, 1997). This involves providing explicit instruction regarding the cues that should be considered when assessing the situation and then presenting examples for the trainee to practice on, together with feedback from a subject matter expert. Despite this, theories of human category learning have not yet been applied to the analysis of the learning process during critical incident training.

In terms of the methodological aims, we introduce process dissociation to the training literature (Jacoby, 1998; Neal et al., 1995). Process dissociation is an experimental technique that allows the user to draw inferences regarding the contribution of underlying cognitive processes to performance on a task. Ford and Kraiger (1995) argued that there is a need to expand the methods and techniques that are used for training needs analysis and evaluation, so that trainers can understand the learning process. Over the past 10 years, we have seen new techniques, such as cognitive task analysis and Pathfinder (Schvaneveldt, 1990) being introduced into the training literature in order to meet this need (Day et al., 2001). Process dissociation, which is described fully in the context of Experiments 1 and 2, is a new technique that falls within this tradition.

The practical aim of this article is achieved by using the results of the process dissociation procedure to train for adaptive performance. The term adaptive performance refers to an individual’s capacity to deal with changing work and exceptional requirements (Hesketh & Neal, 1999; Pulakos, Arad, Donovan, & Plamondon, 2000). To date, most of the research examining the effectiveness of training techniques designed to enhance adaptive performance has focused on generalization. For example, studies have focused on the generalization of knowledge and skills to cases that are more difficult or complex than those encountered during training, or generalization to cases that are seen in a different context (Kozlowski et al., 2001; Schmidt & Bjork, 1992). In the current article, we focus on training techniques designed to enhance a different aspect of adaptive performance. We examine training techniques that may prevent trainees from inappropriately generalizing rules and examples to cases in which they do not apply.

In the following sections, we review current theoretical accounts of category learning and show how they can be used to understand how decision makers learn decision skills. We then introduce the first experiment, in which we use the process dissociation procedure to assess the extent to which trainees rely on two different learning processes: rule learning and exemplar learning. In the second experiment we examine how two aspects of the design and structure of a training program, namely the number of practice examples and the context in which they occur, affect the use of those processes. In doing so, we examine whether these variables can reduce the incidence of trainees inappropriately generalizing from prior examples. In the third experiment we examine whether explicit instruction and task practice can reduce the incidence of trainees inappropriately generalizing a rule.

Category Learning

Historically, there has been a debate over the way in which people learn to perform categorization tasks. One view holds that people learn rules that enable them to perform these tasks (e.g., Bruner, Goodnow, & Austin, 1956). A rule describes a set of criteria by which cases can be evaluated and the way in which those criteria are weighted and combined. In a selection interview, for example, a manager might use intelligence, conscientiousness, and extraversion as the criteria by which she makes a hiring decision. Formal rule-learning models assume that the decision maker evaluates each case on the criteria and assesses where the case falls in relation to a decision boundary on each criterion (Ashby & Gott, 1988; Ashby & Maddox, 1993).

The other view holds that people classify objects and events by considering their similarity to previously seen examples (Brooks, 1978; Medin & Schaffer, 1978). For example, a manager might classify a job applicant as suitable if that person resembles previous applicants that proved to be successful. Exemplar models assume that the decision maker compares the test case with previously seen examples and places the test case in the same category as the examples that it most closely resembles (Kruschke, 1992; Nosofsky, 1988).

Over the past decade, evidence has accumulated suggesting that people use both rules and examples when learning categories (Kulatunga Moruzi, Brooks, & Norman, 2001; Loft, Humphreys, & Neal, 2004; Maddox, Filoteo, Hejl, & Ing, 2004; Roudier & Ratcliff, 2004). For example, experimental studies have demonstrated that people will use prior examples to help them classify an item even when they know the rule and can use it (Allen & Brooks, 1991; Hatala, Norman, & Brooks, 1999). Research has shown that people are more likely to use rules rather than examples when there are a small number of categories, the participants deliberately try to work out the relationship between the dimensions and the category, the feedback enables them to work out this relationship, the rules are easily verbalized, and the items in different categories are easily confused (Justin, Jones, Olsson, & Winman, 2003; Maddox et al., 2004; Roudier & Ratcliff, 2004; Whittlesea, Brooks, & Westcott, 1994). To account for these findings, researchers have developed “dual-process” models, incorporating both rule- and exemplar-learning mechanisms. These dual-process models have successfully accounted for a wide range of phenomena previously taken as evidence for one account or another (Ericsson & Kruschke, 1998; Nosofsky & Palmeri, 1998).

Current evidence from the categorization literature, therefore, suggests that people use a mixture of rule and exemplar strategies when performing categorization tasks. If, as we have argued, recognition-primed decisions are inherently categorical, then it implies that trainees may use both rule and exemplar strategies when learning to make these decisions. To our knowledge, this hypothesis has never been tested before. Although the evidence supporting dual-process models of category learning is relatively strong, the majority of these studies have been carried out in university laboratory settings, using naive participants and highly artificial materials. Both researchers and practitioners have questioned whether findings from the experimental psychology litera-
tute can be applied to the analysis and design of personnel training programs in the field (Latham & Seijts, 1997). Our first goal, therefore, was to assess whether trainees in a real training program use the same learning strategies as participants in controlled laboratory experiments.

Training Design and Structure

If trainees use a mixture of learning strategies during training, then it is important for trainers to understand how the design and structure of a training program influences the use of these strategies, because different strategies have different implications for learning and transfer. Rules enable trainees to generalize to cases that they have never encountered previously. However, in a complex domain, most simple rules will have exceptions (Wittgenstein, 1953). Evidence suggests that if decision makers come to rely on simple rules that work in the majority of situations, then their ability to deal with exceptions to the rule suffers, because they overgeneralize the rule (Ericsson & Krusckbe, 1998; Lewandowsky, Kalish, & Griffiths, 2000; Lewandowsky & Kirsner, 2000). Examples, on the other hand, enable trainees to generalize to cases that are similar to those that they have encountered previously. However, examples do not allow generalization to cases that are dissimilar. For this reason, learning by example will be useful only if the training examples adequately reflect the range of situations that the trainee will encounter in the transfer environment.

The amount of practice that is provided during training is one of the key instructional design parameters that trainees potentially have control over. Practice is known to produce an improvement in performance on a wide variety of tasks (Newell & Rosenbloom, 1981). Mere exposure or experience with a task cannot substitute for practice, because without practice, performance may plateau at a suboptimal level (Ericsson, Krampe, & Tesch-Romer, 1993). However, relatively little research has examined the effect of practice on the development of situation assessment skills or the learning mechanisms that are involved. In one of the few studies addressing this issue, McKinney and Davis (2003) found that the deliberate practice of crisis situations during flight training was associated with better decision making during subsequent crises, although their study did not examine the mechanisms responsible for the effect of practice.

Historically, rule and exemplar category learning models have provided competing explanations for practice effects. Rule models assume that the ability to use a rule improves with practice, through a process of compilation and proceduralization (Anderson & Lebiere, 1998). These models assume that the primary role of practice is to help trainees abstract the rule from a set of examples and to fine-tune the decision criteria. According to a pure rule-learning account, practice should produce an increase in trainees’ use of rules and a reduction in their use of examples. With practice, the number of examples in memory increases, making retrieval faster and easier (Logan, 1988). According to a pure exemplar-learning account, practice should produce an increase in trainees’ use of examples and a reduction in their use of rules.

Dual-process category learning models incorporate both types of mechanisms. They assume that practice allows trainees to fine-tune decision rules and hence that performance will improve in situations where prior examples are not available (Ericsson & Krusckbe, 1998). However, they also assume that when prior examples are available, people will continue to use them, even if they have access to a well-learned rule (Logan, 1988; Rothkopf, Dashen, & Teft, 2002; Vokey & Brooks, 1992; Whittlesea et al., 1994).

Item context is an important design feature that is often overlooked in training. If trainees learn by example, then the contextual features that are associated with the examples will influence their ability to subsequently use those examples. The principles of encoding specificity (Tulving & Thomson, 1973) and transfer-appropriate processing (Morris, Bransford, & Franks, 1977) predict that the retrieval of prior examples from memory will be dependent on the reinstatement of the original encoding context. If the contextual features associated with items in the transfer environment differ from those associated with the training examples, then trainees should find it difficult to retrieve the training examples. It is for this reason that proponents of the concept of situated learning have argued that training examples should be contextualized and that the context of learning should match the context of transfer (Greeno, Moore, & Smith, 1993).

Training for Adaptability

The preceding discussion suggests that trainees can use both rules and examples as a basis for generalizing to new cases. Generalization is often seen as a desirable outcome. Indeed, generalization is one of the major criteria by which people typically evaluate the efficacy of training programs (Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997). However, there are situations where generalization is inappropriate. If trainees are using prior examples to perform a decision task, then we do not want them to apply examples that belong in the wrong category. A trainee might retrieve an example from memory that resembles the case that he or she is currently dealing with, but that belongs in the opposite category. If trainees are using a rule to perform the task, then they need to be able to identify the cases in which the rule does not apply (Felstovich, Spiro, & Coulson, 1997). Training for this type of adaptability presents challenges for trainers, because it is often not possible to give trainees an exhaustive list of the conditions under which the rules and examples do not apply. In a domain such as fire fighting, such a list would be long, difficult to learn, and incomplete. From a practical perspective, therefore, it would be useful to identify some of the factors that reduce the incidence of trainees inappropriately generalizing from rules and examples.

A pure rule-learning account suggests that practice should reduce the incidence of inappropriate generalization from examples. As noted above, rule-learning models assume that practice produces a shift from the use of examples to the use of rules. As trainees begin to use a rule more effectively, the likelihood of applying the wrong example should decrease. However, dual-process models of category learning predict that trainees will continue to use prior examples when they are available. As a result, practice may not be effective. Dual-process models suggest that context is likely to be more effective. Trainees will be less likely to generalize inappropriately by applying the wrong exam-
ple if the contextual features of the transfer item do not match those of the practice examples.

Feltovich et al. (1997) argued that trainees are more likely to recognize exceptions to a general rule if they adopt a personal epistemology, or belief structure, that emphasizes the complexity of the domain and acknowledges that simple rules and procedures do not apply to all cases. This argument suggests two strategies for developing this kind of adaptability. These strategies correspond to the two learning processes examined in the earlier experiments. The first strategy involves encouraging trainees to develop a more complex rule that has fewer exceptions. One way to do this may be through explicit instruction. Simply instructing trainees about the complexity of the domain and the types of exceptions that are encountered may encourage them to attend to a larger number of dimensions when learning and to develop rules that are more complex (Feltovich et al., 1997). This should enhance performance in situations where the simple rule does not apply.

The second strategy involves learning from examples. Exceptions to the rule can be illustrated using examples. Examples may be useful for two reasons. First, trainees may store these examples in memory and may subsequently recognize these types of exceptions when they encounter them again (Jones & Endsley, 2000). Second, incorporating examples of exceptions into practice may encourage effortful processing, particularly if the exceptions are surprising and trainees make errors on them (Hesketh, 1997; Ivancic & Hesketh, 2000). This may encourage trainees to think more critically (Cohen et al., 2000), leading them to an awareness of the limitations to the generality of the rule that they have learned and encouraging them to develop a rule that is more complex (Feltovich et al., 1997). Again, this should enhance performance in situations where the simple rule does not apply.

**Experiment 1**

The goal of Experiment 1 was to examine cognitive processes used by trainees when learning decision skills. As recommended by Cohen et al. (2000) and Klein (1997), we used a critical incident method for training decision skills, in conjunction with explicit instruction. The examples, which were generated by subject matter experts, were constructed so that they could be correctly classified using a relatively simple rule. The trainees had to make decisions and were then given feedback from one of the experts. The feedback illustrated how they had used the rule in that situation. We provided explicit instruction in the form of a lecture before presenting the examples. Although there has been concern expressed about the role of the lecture, a recent meta-analysis has demonstrated that lectures can enhance learning (Arthur, Bennett, Edens, & Bell, 2003). Explicit instruction techniques such as the lecture are thought to provide a schema that directs trainees’ attention to the critical features of the practice examples and enables them to use the feedback more effectively (Cohen et al., 2000). The rule that could be used to correctly classify the training examples is described in the Method section, below.

The specific aim of Experiment 1 was to assess the extent to which trainees use rules and examples when learning to make tactical decisions. Our training program focused on teaching firefighters to assess whether it is safe to enter a building. Unfortunately, cognitive processes are not directly observable, and there are relatively few techniques available for drawing inferences about cognitive processes. The critical decision method (Klein, Calderwood, & MacGregor, 1989) is perhaps the best known technique within the applied psychology literature. In this technique, the analyst asks the decision maker to recall a difficult or challenging incident and then analyzes the verbal protocol for evidence of different decision strategies (e.g., the cues that were attended to, the number of decision alternatives that were considered). However, this technique is retrospective and subjective and can only reveal processes that are available to consciousness and can be recollected.

Process dissociation, on the other hand, is a comparatively objective technique. It is not dependent on the analysis of verbal protocols and allows the analyst to draw inferences from patterns of responses about the operation of automatic processes, as well as controlled or conscious processes. Originally, process dissociation was developed to allow researchers to draw inferences regarding the contribution of intentional and automatic processes to recognition memory (Jacoby, 1991). Since then, the technique has been used for a variety of tasks, including category learning (Neal et al., 1995).

The logic of process dissociation is that the contribution of two hypothetical processes can be estimated by comparing performance in a condition in which the processes act together (an “inclusion” condition) with performance in a condition in which they act in opposition (an “exclusion” condition). We created inclusion and exclusion conditions by varying the relationship between training and test items. Table 1 contains some examples of training and test items. As can be seen in Table 1, each training example had one matching test item. These test items were similar to their matching training examples on all but one dimension, which was switched to its opposite value. For example, Item 4b is very similar to Item 4a except for one feature—in this case, the building structure is different (brick wall vs. wooden floors and ceilings). Similarly, Item 6b is very similar to Item 6a except for one feature—such as the intensity of the fire is likely to be different because the contents of the building differ (garden furniture vs. paints and varnishes).

The switching of these features resulted in the analogous test items either staying in the same category as their matching training example (inclusion) or shifting to the opposite category (exclusion). Item 6b is an inclusion item, because changing the contents from garden furniture to paints did not change the category that the item belongs in. In this case, both the training and test items were unsafe to enter. In the inclusion conditions, participants would arrive at the correct response if they used the rule or if they classified the test item by analogy to its matching training example. Item 4b is an exclusion item, because changing the structure from double brick to brick veneer does change the category that the item belongs in. If the trainees used the rule, they would know that this building is not safe to enter. However, if they classified this test item by analogy to its matching training item, then they would mistakenly think that it is safe to enter. Thus, in the exclusion conditions, participants would arrive at different responses, depending on which process they used. We also presented novel test items, which were not matched to specific training examples. Item 12 is an example of a novel item. The novel items assess participants’ ability to generalize to new cases when they cannot use prior examples.
The procedure described above allows us to draw inferences regarding the use of rules and examples. If our assumptions are correct, then performance on the novel items provides an indication of trainees’ ability to use a rule when there are no matching examples available. Performance on the inclusion and exclusion items, on the other hand, provides an indication of the extent to which trainees use prior examples. If the trainees use prior examples to help them classify the test items, then performance on the inclusion items should be enhanced relative to the novel items whereas performance on the exclusion items should be impaired relative to the novel items. In other words, one would expect accuracy to be highest on the inclusion items and lowest on the exclusion items, with performance on the novel items falling in between. We would also expect a similar pattern to emerge with decision speed and confidence. Given the arguments presented previously, we expected that the trainees would use both rules and examples when learning situation assessment skills. Our hypotheses were, therefore, as follows:

**Hypothesis 1:** Performance on the novel items will be significantly better than chance.

**Hypothesis 2:** Performance on the inclusion items will be better than performance on the novel items.

**Hypothesis 3:** Performance on the exclusion items will be worse than performance on the novel items.

### Table 1
**Sample Training Items, Together With Their Matching Inclusion and Exclusion Test Items, and a Novel Test Item**

<table>
<thead>
<tr>
<th>Training items</th>
<th>Test items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 4a (Safe to enter) A 000 call has been received at 2 in the afternoon to a factory in Liverpool. The building is double brick with concrete floors [building structure: resistant to fire] and is used for the production of highly flammable and potentially unstable chemical products [intensity of fire: high]. You have just arrived at the fire and have been informed that all workers have been accounted for. At this stage, the fire is burning in the upstairs test laboratory, but the main production area is situated on the lower floor [fire size: small]. A 275-mm main is located on the street, with a hydrant located on the footpath near the front entrance.</td>
<td>Item 4b (Exclusion item, unsafe to enter) A 000 call has been received at 7 a.m. to a factory in Liverpool. The building is brick veneer with polished floors and ceilings [building structure: not resistant to fire] and is used for the production of highly flammable and potentially unstable cleaning products [intensity of fire: high]. You have just arrived at the fire and have been informed that all workers have been accounted for. At this stage, the fire is in the upstairs quality control laboratory, but the main production area is situated on the lower floor [fire size: small]. The fire is directly impinging on a substantial area of both the floor and roof framing on the first floor. A 300-mm main is located on the street, with a hydrant located next door.</td>
</tr>
<tr>
<td>Item 6a (Unsafe to enter) A 000 call has been received at 2 a.m. to a factory in Alexandria. The building is constructed in timber and corrugated iron [building structure: not resistant to fire] and is used for the production and storage of garden furniture [intensity of fire: low]. You have just arrived at the fire and have been informed that all night workers have been accounted for. At this stage, the fire has spread throughout the factory floor [fire size: large]. A 300-mm main is located on the street, with a hydrant located opposite the back entrance.</td>
<td>Item 6b (Inclusion item, unsafe to enter) A 000 call has been received at 8 p.m. to a factory in Alexandria. The building is constructed in double brick [building structure: resistant to fire] and is used for offices for various private businesses. You have just arrived at the fire and have been informed that all workers have been accounted for. At this stage, the fire, which started in the varnish production area, has spread through most of the factory. Varnish products in both storage and production are now alight, and have spread to the timber framing [fire size: large]. A 350-mm main is located on the street, with a hydrant located on the roadway.</td>
</tr>
<tr>
<td>Item 12 (Novel item, safe to enter) A 000 call has been received at 3 p.m. time to a multistory building in Kings Cross. The building is constructed in double brick [building structure: resistant to fire] and is used as offices for various private businesses. You have just arrived at the fire and have been informed that all workers have been accounted for. At this stage, the fire is in one upstairs office [fire size: small] and fueled by a spill of printing chemicals [intensity of fire: high]. A 275-mm main is located on the street, with a hydrant located next door.</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Text in italics indicates status of relevant dimensions and was not read out. A “000” call is an emergency call.

### Method

**Research setting.** The study was carried out in the context of an introductory training program for urban firefighters. Fire fighting was chosen because it requires decisions to be made on the basis of incomplete information, in a dynamic environment, often under time pressure. Consequently, situation assessment is one of the most important skills that firefighters need to develop (Klein, Calderwood, & Clinton-Cirocco, 1986). Specifically, firefighters need to be able to assess whether a building is safe to enter. This is not a straightforward decision. Although it is possible to generate relatively simple rules of thumb that describe the dimensions that should be considered, it is difficult to describe the ways in which each of the dimensions should be weighted and combined. Furthermore, much of the information is given verbally by others and is imprecise and requires interpretation.

**Participants.** There were 35 participants, all of whom were new recruits to the New South Wales (NSW) Fire Brigades. All participants had undergone 3 weeks of training prior to participating in this study. The 3-week training program focused primarily on manual skills and did not include material relating to fire fighting tactics, fire behavior, structural integrity, or building collapse. The module incorporated into this study was presented at the end of the 3-week training program. It was designed to introduce trainees to tactical decisions to provide a broader context for their preliminary training. Participation was voluntary, and participants were free to withdraw at any time. Thirty-three participants were male and 2 were female, with an average age of 30.9 years. Previous fire fighting experience ranged from 0 to 12 years, with an average of 3.1 years. The participants with prior experience had typically obtained that experience...
working as volunteers for the rural fire service, attending bush fires. As a result, their experience with fires in urban contexts was limited.

Materials. A lecture was written to provide the trainees with information on fire fighting tactics, fire behavior, structural integrity, and building collapse. The lecture explained the factors that influence fire intensity (primarily type and quantity of fuel) and listed the types of dimensions that should be considered when assessing the probability of building collapse. These included the size of the fire, the intensity of the fire, building structure, the apparent age of the building, and the length of time the fire had been burning. The lecture lasted 15 min.

There were 8 training examples and 12 test examples. Each example consisted of a photograph of a structural fire and an accompanying auditory script. The photos were displayed on a 14-in. computer monitor and occupied the entire screen using an 800-by-600 pixel resolution. The auditory scripts were played over headphones. The scripts were generated in collaboration with a subject matter expert to ensure that they were realistic. The scripts described the state of the fire as would be observed on first arrival to the fire. Each script contained information regarding three dimensions that influence the likelihood of building collapse (building structure, size of the fire, and intensity of the fire) along with a range of additional features that a fire commander would note in his or her initial “size-up” of the situation but that do not influence the specific decision to enter the building or stay outside. Because these additional features do not determine the categorical membership of the items, we refer to them as “surface features.” The surface features varied from item to item and could include the suburb that the building was in, occupancy type, building use, time of day, hydrant location, and size of the water main.

As is typically the case in categorization research, the dimensions were binary. Thus, each dimension could take on one of two values. For example, the building structure could be resistant to the effects of fire or not resistant to the effects of fire; the fire could be large or small; and the intensity of the fire could be high or low. The Appendix shows the values of each dimension for every item. This table defines a category structure similar to that used by Allen and Brooks (1991). The two categories consisted of buildings that were safe to enter and buildings that were not safe to enter. The category structure was designed so that each relevant dimension predicted category membership on 75% of occasions, and membership could be perfectly predicted using a three-feature additive rule. According to this rule, a building was safe to enter if at least two of the following conditions were met: The structure was resistant to the effects of fire, the fire was small, and the fire was not intense. If not, then the building was unsafe.

As explained earlier, each training example had one matching test item, which could be either an inclusion or an exclusion item. There were four inclusion items and four exclusion items. As can be seen in Table 1, the surface features of the inclusion and exclusion items were very similar to those in their matching training examples. However, the value for one of the relevant dimensions was switched. For the exclusion items, the switching of the relevant dimension changed the category that the item belonged in. For the inclusion items, the switching of the relevant dimension did not change the category that the item belonged in. In addition, four novel test items were created, which did not resemble any training items. Unlike the inclusion and exclusion items, which had matching surface features, the novel items had unique surface features. For example, the novel item shown in Table 1 occurred in a different suburb and had a unique occupancy type. Furthermore, the specific type of fuel (“a spill of printing chemicals”) was unique to that item.

Procedure. The participants were first given a prerecorded lecture, which was presented on the computer. Each section of the lecture was followed by a set of written multiple-choice questions to ensure that the trainees had attended to and understood the lecture. There were 16 questions in total. Example questions included “How does fire spread to external exposures?” (Answer: Fire spreads to external exposures through projected radiant heat); “What is the aim of an internal offensive attack?” (Answer: An internal offensive attack is aimed at achieving rapid containment to reduce the damage from heat, smoke, and water). After finishing the questions, the participants clicked the mouse to reveal the correct answer on the screen. To encourage the trainees to adopt a learning orientation rather than a performance orientation (Button, Mathieu, & Zajac, 1996), we did not collect or mark the answer sheets. The purpose of the questions was simply to encourage the trainees to attend to the information in the lecture and think about it.

Following the lecture, participants were given the training examples. The training examples were shown in a random order. For each example, the photograph of the fire was first presented on the screen, and the auditory script was then played. After the script had finished, the trainees were asked to make a series of decisions about the fire. The key decision that we were interested in was whether they would enter the building. Each question was presented in multiple-choice format. Response time data were collected automatically.

The first two questions were filler questions, designed to get the participants thinking about the types of issues that they would consider in the field. These questions were (a) “What exposures would you expect at this type of fire?” and (b) “Where will you try to contain the fire?”

The third question assessed whether the trainee would enter the building. This was phrased in terms of attack strategy (“What initial attack would you employ?”). Four options were presented: (a) “Send in crews with hoses to fight fire”; (b) “Send in crews with hoses to protect internal exposures”; (c) “Position crews outside with hoses to fight fire”; and (d) “Position crews outside with hoses to protect external exposures.” Options a and b involve entering the building, whereas Options c and d involve staying outside. The reason for embedding the categorical decision within a multiple-choice test was that we wanted to avoid asking participants to make an explicit categorization response. Evidence suggests that people may perform categorization tasks differently if asked to make an explicit categorization decision (“Is this building safe to enter?”) than if asked to perform a task that requires an implicit categorization decision (“What initial attack will you employ?”) (Markman & Ross, 2003).

The fourth question assessed participants’ confidence in the appropriateness of their attack strategy. Five response options were presented for this question, from not at all confident to very confident. Each question was answered by mouse-clicking a button next to the answer.

After answering all four questions, participants were shown the photograph of the fire a second time, with a voice-over describing how the fire was resolved. This voice-over described what the exposures were, where the fire was contained, how the fire should have been attacked, and the reasons why such an attack was used.

The test phase was presented after the training phase had been completed. The same procedure was used as in the training phase, except that feedback was not given describing how the fire was resolved. The order of presentation of the 12 test items was randomized for each participant.

Analysis. The test data were analyzed using a series of repeated measures analyses of variance (ANOVA)s. Planned contrasts were run, comparing performance on the inclusion items with the novel items, and the novel items with the exclusion items. The primary dependent variable was classification accuracy, which was based on participants’ responses to Question 3. As is commonly the case in decision-making research, we used reaction time and confidence as secondary dependent variables. Examining accuracy in relation to reaction time and confidence allows us to draw stronger inferences regarding the nature of the underlying decision process.

Results

Table 2 shows the means, standard errors, and correlations among the variables. As would be expected, the reaction time variables are correlated with each other, as are the confidence ratings. Confidence ratings for inclusion items are also correlated with accuracy for those items. Thus, there appear to be individual
differences in the speed and confidence of decisions; however, there is no consistent evidence of a speed-accuracy trade-off that might confound the results.

Overall, the trainees performed relatively well on the inclusion and novel items. As predicted, performance on the novel items was significantly better than chance (Hypothesis 1), \( t(34) = 12.154, p < .001, d = 2.05 \). Furthermore, a higher proportion of the inclusion items were answered correctly than the novel items (Hypothesis 2), \( F(1, 34) = 30.90, p < .001, \eta^2 = .48 \), and a lower proportion of the exclusion items were answered correctly than the novel items (Hypothesis 3), \( F(1, 34) = 38.18, p < .001, \eta^2 = .53 \). Reaction times were faster for the inclusion items than for the novel items, \( F(1, 34) = 7.38, p < .01, \eta^2 = .18 \). However, there was no difference between the reaction times for novel items and exclusion items. Confidence was higher for the inclusion items than for the novel items, \( F(1, 34) = 12.99, p < .01, \eta^2 = .28 \), and was lower for the exclusion items than for the novel items, \( F(1, 34) = 5.04, p < .05, \eta^2 = .13 \).

**Discussion**

The results from Experiment 1 support all three hypotheses. First, as demonstrated by the fact that performance on novel items was better than chance, the trainees were able to generalize to items that were not matched to specific training examples. According to the logic of the process dissociation procedure, the ability to transfer to novel items suggests that the trainees had learned to use a general rule. There were no matching training items available to help them classify these novel items.

Second, participants performed better on the inclusion items than on the novel items and were able to answer these items more quickly and with greater confidence. According to the logic of the processes dissociation procedure, this effect allows us to infer that the trainees were using prior examples. The key difference between the inclusion items and the novel items was that the inclusion items closely resembled a training example that belonged in the same category, whereas the novel items did not. If the trainees were only using rules, then we would not have observed any differences between inclusion and novel items in accuracy, reaction time, or confidence.

Third, the participants performed worse on the exclusion items and reported feeling less confident about their answers. This effect supports the inferences drawn from the analysis of the inclusion items. The exclusion items closely resembled a training example that belonged in the opposite category. In this case, the trainees appear to have generalized inappropriately from prior examples, resulting in poor performance in this condition. It is interesting to note that the participants were less confident about their decisions for the exclusion items than for the novel items, but reaction times were not affected. This suggests that participants were aware that their exclusion decisions might not be correct but did not attempt to compensate by giving themselves more time to make the decision.

In summary, the process dissociation procedure has allowed us to draw inferences regarding the extent to which trainees used rules and examples when learning from critical incidents. When the rule-learning and exemplar-learning processes were placed in conjunction with each other, performance was better than when the rule-learning process acted in isolation. On the other hand, when the rule-learning and exemplar-learning processes were placed in opposition to each other, performance was worse than when the rule-learning process acted in isolation. The results are therefore consistent with the prediction that trainees would use both rules and examples when learning to make tactical decisions.

**Experiment 2**

The aim of Experiment 2 was to examine the effects of task practice and context on the use of rules and examples. Doing so allowed us to test the validity of the process dissociation inferences drawn in Experiment 1. If the inferences regarding the use of rules and examples are valid, then these processes should respond differentially to task manipulations that specifically target each learning process. As will be seen below, we predicted that task practice and context would have differential effects on each learning process.

As noted earlier, dual-process models of categorization predict that practice will produce an improvement in trainees’ ability to use a rule when prior examples are not available but that trainees will continue to use prior examples when they are available. We examined the effects of practice in Experiment 2 by inserting an extra test phase halfway through the training phase. This allowed us to assess whether the use of rules and examples changed as the number of examples increased. As in Experiment 1, performance on the novel items provides an indicator of trainees’ ability to use a rule when prior examples are not available, while the difference...
in performance on the inclusion items and exclusion items provides an indicator of the extent to which trainees use examples when they are available. Our first two hypotheses, therefore, were as follows:

**Hypothesis 1:** Performance on the novel items will improve from Phase 1 to Phase 2.

**Hypothesis 2:** The differences in performance on the inclusion items and exclusion items will remain stable over practice.

Thus, in Experiment 2, we were expecting to observe a single-task dissociation. That is, we were expecting to find that trainees’ ability to use a rule would improve with practice, even though they would continue to use examples when they were available. Single-task dissociations occur when one variable has a significant effect on one type of task (or measure) but has no effect on another task (or measure). This pattern of data is often taken as evidence that separate processes underlie performance on the two tasks or measures (Maddox et al., 2004). In this case, the pattern described in Hypotheses 1 and 2 would support the claim that the use of rules during training is independent of the use of examples. However, a single-task dissociation by itself provides relatively weak evidence for the existence of separate processes, because the absence of a significant effect of the experimental variable on one of the tasks could simply reflect the fact that the measures have differential sensitivity. For this reason, we incorporated a second experimental manipulation designed to have the opposite effect, producing a double-task dissociation. Double-task dissociations provide stronger evidence for the existence of separate underlying processes, because they cannot be explained on the basis of differential sensitivity.

We expected that item context would have the opposite effect to that of task practice. As noted earlier, the principles of encoding specificity and transfer-appropriate processing predict that the effects of prior examples would be stronger when the contextual features of matching training and test items were similar. We manipulated context by varying the photographs of the buildings that were paired with the scripts in the test phase. Half of the inclusion items and half of the exclusion items were in a building that was very similar to the building that the matching training fire was in. The remaining inclusion and exclusion items were in a building that was different from the building that the matching training fire was in. Our third hypothesis was as follows:

**Hypothesis 3:** The differences in performance on the inclusion items and exclusion items will be larger when the contextual features of the training and test items are matched.

### Method

**Design.** To manipulate practice and context, we split the item pool in half and counterbalanced the order of presentation of the two pools. One set of training items had its matching inclusion and exclusion items shown in similar buildings. This set of training and test items is called the “same-context set.” The other set of training items had its matching inclusion and exclusion items shown in different buildings. This set of training items is called the “different-context set.”

The experiment was divided into two phases. In the first phase, half of the participants were given the same-context training items followed by the same-context test items. In the second phase, these participants were given the different-context training items followed by both the same-context and different-context test items. The other group received the reverse order of presentation (see Table 3). This design allowed us to examine the way in which performance on the inclusion items, novel items, and exclusion items changes from Phase 1 to Phase 2 (Hypotheses 1 and 2), and as a function of context (Hypothesis 3), while controlling for potential differences between items in the two pools.

**Participants.** There were 36 participants, all of whom were new recruits to the NSW Fire Brigades. All participants had undergone 3 weeks of training but had not been educated on aspects of fire management during their training at the time of taking part in the experiment. Thirty-four were male and 2 were female, with an average age of 31.4 years. Previous fire fighting experience ranged from 0 to 22 years, with an average of 5 years. None of the participants in Experiment 2 were involved in Experiment 1.

**Materials.** The lecture and the training and test items were the same as those used in Experiment 1. Four of the training items were assigned to the same-context set, and the other four were assigned to the different-context set. Each test set had two inclusion items that matched items in its corresponding training set, two exclusion items that matched items in its corresponding training set, and two novel items. Context was manipulated by varying the photographs that were paired with the inclusion and exclusion items. The photographs that were paired with the same-context inclusion and exclusion items were selected so that the buildings were closely matched in terms of age, construction, and appearance, and the fires appeared similar from the outside. Because of the difficulty of finding photographs of fires in buildings with a specific type of construction, we had to use photographs from the same fire on occasion. The photographs were taken from different angles and had different people and equipment in them. As a result, they looked similar but not identical. The photographs that were paired with the different-context inclusion and exclusion items were taken from different fires. Although the buildings had the same occupancy type, they differed in terms of age, construction, and appearance.

**Procedure.** The participants were given the lecture and were then given the first training set (same context or different context). After the participants had completed the first training set, they were given the relevant set of test items. Specifically, the group given the same-context training set was given the same-context test set, and the group given the different-context training set was given the different-context test set. After completing the first test phase, the participants were given the second

### Table 3

**Design of Experiment 2**

<table>
<thead>
<tr>
<th>Group</th>
<th>Phase 1 Training Set</th>
<th>Phase 1 Test Set</th>
<th>Phase 2 Training Set</th>
<th>Phase 2 Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Same-context training set</td>
<td>Same-context test set</td>
<td>Different-context training set</td>
<td>Same-context and different-context test sets</td>
</tr>
<tr>
<td>Group 2</td>
<td>Different-context training set</td>
<td>Different-context test set</td>
<td>Same-context training set</td>
<td>Same-context and Different-context test sets</td>
</tr>
</tbody>
</table>
training set. The second test phase was then presented. Both sets of test items (same and different context) were presented in the second test phase.

**Analysis.** The test data were analyzed using a series of repeated measures ANOVAs. Planned contrasts were run, assessing the effects of practice on the novel items and the effects of practice and context on the inclusion and exclusion items. The dependent variables were classification accuracy, classification reaction time, and confidence.

**Results.**

Table 4 shows the means, standard errors, and correlations among the variables used in Experiment 2. As in Experiment 1, the confidence measures are correlated with each other. However, compared with the results in Experiment 1, the reaction time measures are not as strongly correlated with each other. Instead, there is a differential pattern of correlations among the accuracy, confidence, and reaction time measures for the novel and exclusion items. For the novel items, accurate decisions were made quickly and confidently. Accuracy is positively correlated with confidence, whereas both accuracy and confidence are negatively correlated with reaction time. For exclusion items, inaccurate decisions were made quickly and confidently. Accuracy was positively correlated with reaction time and negatively correlated with confidence. This pattern of data is consistent with the argument that it is memory retrieval that is causing the errors on the exclusion items. The prior example is likely to come to mind relatively quickly and cause an error if the trainee does not deliberately check his or her answer using the rule.

**Practice.** Figure 1 shows the effects of task practice on performance for the inclusion items, novel items, and exclusion items. The first analysis tested the prediction that performance on the novel items would improve with practice (Hypothesis 1). As expected, there was a significant improvement in accuracy for the novel items when the contexts matched (inclusion: $M = 0.95, SE = 0.02$; exclusion: $M = 0.55, SE = 0.04$), $F(1, 34) = 181.511, p < .05, \eta^2 = .84$, and the difference between inclusion and exclusion items did not change with practice, $F(1, 34) = 0.06, p > .05, \eta^2 = .00$. There were no interactions between item type (inclusion vs. exclusion) and practice for confidence or reaction time, either.

**Context.** Figure 2 shows the effects of context on performance for the inclusion and exclusion items. Novel items are not shown in this figure, because by definition, they could not have a matching context. Our third analysis tested the prediction that differences in performance on the inclusion and exclusion items would be larger when the contextual features of the training and test items were matched (Hypothesis 3). As expected, the difference between inclusion and exclusion items was stronger when the contexts matched, $F(1, 34) = 97.33, p > .05, \eta^2 = .74$. An interaction between item type and context also emerged in the context and confidence data. Reaction times were faster for inclusion items than for exclusion items when the contexts matched (inclusion: $M = 3.6 s, SE = 0.2 s$; exclusion: $M = 4.8 s, SE = 0.4 s$) but not when the contexts were different (inclusion: $M = 4.4 s, SE = 0.4 s$; exclusion: $M = 3.8 s, SE = 0.3 s$), $F(1, 34) = 9.43, p < .05, \eta^2 = .22$. Confidence was higher for inclusion items than for exclusion items when the contexts matched (inclusion: $M = 4.4, SE = 0.1$; exclusion: $M = 3.7, SE = 0.1$) but not when the contexts were different (inclusion: $M = 4.2, SE = 0.1$; exclusion: $M = 4.0, SE = 0.1$), $F(1, 34) = 14.78, p < .05, \eta^2 = .30$.

**Discussion**

The results from Experiment 2 supported all three hypotheses. Practice was expected to produce an improvement in trainees’ ability to use the rule, without having any effect on their tendency to use examples. Our indicator of rule use was trainees’ ability to generalize to novel problems. The results show that increasing the number of examples from four to eight did facilitate generalization to novel problems. Our indicator of the use of examples was the difference in performance on the inclusion and exclusion items. The improvement in performance on novel items was not accompanied by any corresponding reduction in the difference between inclusion and exclusion items. These findings suggest that trainees continued to use prior examples even though their ability to use the rule had improved. Practice by itself, therefore, does not appear to be sufficient to prevent trainees from generalizing inappropriately from prior examples.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$M$</th>
<th>$SE$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusion accuracy</td>
<td>0.96</td>
<td>0.01</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Novel accuracy</td>
<td>0.75</td>
<td>0.03</td>
<td>-.04</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Exclusion accuracy</td>
<td>0.56</td>
<td>0.03</td>
<td>.17</td>
<td>-.08</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Inclusion reaction time</td>
<td>4.05</td>
<td>0.23</td>
<td>.18</td>
<td>-.07</td>
<td>.21</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Novel reaction time</td>
<td>4.37</td>
<td>0.30</td>
<td>.19</td>
<td>-.39*</td>
<td>.46*</td>
<td>.52*</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Exclusion reaction time</td>
<td>4.39</td>
<td>0.31</td>
<td>.08</td>
<td>.15</td>
<td>.15</td>
<td>-.28</td>
<td>-.11</td>
<td>-.23</td>
<td>.05</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Inclusion confidence</td>
<td>3.88</td>
<td>0.12</td>
<td>.00</td>
<td>.53*</td>
<td>-.28</td>
<td>-.19</td>
<td>-.43*</td>
<td>-.05</td>
<td>.79*</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Exclusion confidence</td>
<td>3.82</td>
<td>0.13</td>
<td>-1.7</td>
<td>.23</td>
<td>-.34*</td>
<td>-.20</td>
<td>-.39*</td>
<td>-.22</td>
<td>.77*</td>
<td>.82*</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note.* $N = 36$.  
* $p < .05$. 

---

[Table 4] Means, Standard Errors, and Correlations Among Variables in Experiment 2

---

[Figure 1] The effects of task practice on performance for the inclusion items, novel items, and exclusion items. The first analysis tested the prediction that performance on the novel items would improve with practice (Hypothesis 1). As expected, there was a significant improvement in accuracy for the novel items when the contexts matched (inclusion: $M = 0.95, SE = 0.02$; exclusion: $M = 0.55, SE = 0.04$), $F(1, 34) = 181.511, p < .05, \eta^2 = .84$, and the difference between inclusion and exclusion items did not change with practice, $F(1, 34) = 0.06, p > .05, \eta^2 = .00$. There were no interactions between item type (inclusion vs. exclusion) and practice for confidence or reaction time, either.

[Figure 2] The effects of context on performance for the inclusion and exclusion items. Novel items are not shown in this figure, because by definition, they could not have a matching context. Our third analysis tested the prediction that differences in performance on the inclusion and exclusion items would be larger when the contextual features of the training and test items were matched (Hypothesis 3). As expected, the difference between inclusion and exclusion items was stronger when the contexts matched, $F(1, 34) = 97.33, p > .05, \eta^2 = .74$. An interaction between item type and context also emerged in the context and confidence data. Reaction times were faster for inclusion items than for exclusion items when the contexts matched (inclusion: $M = 3.6 s, SE = 0.2 s$; exclusion: $M = 4.8 s, SE = 0.4 s$) but not when the contexts were different (inclusion: $M = 4.4 s, SE = 0.4 s$; exclusion: $M = 3.8 s, SE = 0.3 s$), $F(1, 34) = 9.43, p < .05, \eta^2 = .22$. Confidence was higher for inclusion items than for exclusion items when the contexts matched (inclusion: $M = 4.4, SE = 0.1$; exclusion: $M = 3.7, SE = 0.1$) but not when the contexts were different (inclusion: $M = 4.2, SE = 0.1$; exclusion: $M = 4.0, SE = 0.1$), $F(1, 34) = 14.78, p < .05, \eta^2 = .30$. 

**Discussion**

The results from Experiment 2 supported all three hypotheses. Practice was expected to produce an improvement in trainees’ ability to use the rule, without having any effect on their tendency to use examples. Our indicator of rule use was trainees’ ability to generalize to novel problems. The results show that increasing the number of examples from four to eight did facilitate generalization to novel problems. Our indicator of the use of examples was the difference in performance on the inclusion and exclusion items. The improvement in performance on novel items was not accompanied by any corresponding reduction in the difference between inclusion and exclusion items. These findings suggest that trainees continued to use prior examples even though their ability to use the rule had improved. Practice by itself, therefore, does not appear to be sufficient to prevent trainees from generalizing inappropriately from prior examples.
Context was expected to have the opposite effect to that of practice. The principles of encoding specificity and transfer-appropriate processing predict that the retrieval of prior examples will be enhanced when the contextual features of the test item are similar to those associated with the training example. The results supported this prediction. The differences between the inclusion and exclusion items were larger when the contextual features associated with the matching test items were similar to those used during training. This suggests that the retrieval and use of prior examples was dependent on contextual overlap.

It is also worth noting that the pattern of correlations among the accuracy, reaction time, and confidence measures supports the assumptions underlying the process dissociation procedure. The memory retrieval process that produces the errors on the exclusion items should be relatively fast (Jacoby, 1991). The reason why this pattern did not emerge in Experiment 1 may simply be that we used more items in the current experiment and hence had more reliable measures for analyses of individual differences.

From a theoretical and practical perspective, the results of Experiment 2 suggest that the trainees’ learning strategies can be shaped by specific aspects of training structure and design, namely the amount of practice allowed and the range of contexts used. Specifically, the results suggest that if trainers wish to enhance their trainees’ ability to use general rules, then this may be achieved by increasing the number of practice examples. However, if trainers wish to enhance their trainees’ ability to use specific examples, then this may be achieved by ensuring that the contextual features of the training examples match those encountered in the transfer environment. This, of course, has both costs and benefits. As can be seen in both experiments, there is the risk that trainees will generalize inappropriately from prior examples.

From a methodological perspective, the results from Experiment 2 support the inferences drawn in Experiment 1. The indicators of rule use and example use responded differentially to experimental manipulations designed to selectively target each process. This pattern of results of Experiment 2 represents a double dissociation and supports the claim that the indicators used in these experiments do map onto their hypothesized underlying processes. We believe that the procedure used in Experiment 2 represents a useful way for training researchers to check the validity of the inferences that they draw from the process dissociation procedure.

### Experiment 3

The aim of Experiment 3 was to examine the effectiveness of training strategies designed to prevent trainees from generalizing a simple rule to cases in which that rule does not apply. As outlined in the introduction, one of the training strategies that we used was explicit instruction regarding exceptions to the rule. The other strategy that we examined involved the use of examples to illustrate exceptions to the rule. In the test phase, we presented a series of cases that represented exceptions to the rule. Some of the exceptions were of the same general type as encountered during training (near exceptions), whereas others were completely new (far exceptions). The far exceptions provide the most stringent test of the effectiveness of these training strategies, because they require the trainees to notice the existence of a unique set of circumstances that they have never encountered previously. The far exceptions require the trainees to use their discretion and ignore a rule that they have just learned. Our hypotheses were as follows:

**Hypothesis 1:** Providing trainees with explicit instruction regarding exceptions to a simple rule will enhance performance on near and far exceptions.

**Hypothesis 2:** Providing trainees with examples that illustrate exceptions to a simple rule will enhance their performance on near and far exceptions.

### Method

A $2 \times 2 \times (3)$ design was used. The first between-groups factor was instruction. There were two types of instruction: routine instruction and exceptions instruction. As in the previous experiments, all trainees were given a lecture that explained some of the general principles of fire behavior and management. However, half of the participants (the exceptions instruction group) received additional information in this lecture. The
exceptions instruction group were alerted to the fact that factors such as construction, fire size, and fire intensity interact in complex ways (e.g., that steelwork not encased in concrete is vulnerable to the effects of high-intensity fires). They were also provided with information regarding the types of cues that could indicate possible building collapse (e.g., sagging floors and ceilings, cracked or dropping arches over windows, falling cornices and ceiling material) and the factors that should be considered when carrying out search and rescue (e.g., immediacy of threat, number of people to be rescued, and location of people in the structure). The exceptions instruction group, therefore, were given information regarding some of the conditions under which the general rule (based on an additive combination of structure, size, and intensity) does not apply, at least in its simple form. This additional information added one and a quarter minutes to the lecture.

The second between-groups factor was practice. There were two sets of practice conditions: routine practice and exceptions practice. As in the previous experiments, all of the participants received eight practice examples. In the routine practice condition, all eight examples could be answered correctly using the three-feature additive rule based on structure, size, and intensity. In the exceptions practice condition, two of the examples were modified in order to illustrate exceptions to this rule. The first exception involved a building that would normally be considered safe to enter, because although the fire was large, it was of low intensity and the construction was of cement and steel. However, the type of construction used (concrete tilt slab) created a risk of the walls collapsing inward, making it dangerous to enter. The second exception involved a building that would normally be considered dangerous to enter, because the fire was large and the construction was potentially unstable. However, a search and rescue operation was required because the fire was in a foster home, and all of the residents had not been accounted for.

The within-group factor was item type. There were 16 test items, consisting of three different types: routine (8 items), near exceptions (4 items), and far exceptions (4 items). The routine items could be answered correctly using the rule. The near exceptions involved the same types of situations that were previously shown to the exceptions practice group (imminent collapse and search and rescue). The far exceptions involved new situations (e.g., a building that should not be entered because it contains potentially explosive materials or a building that needs to be entered in order to protect priceless artworks). None of the participants had encountered these situations in the lecture or during the practice phase. Furthermore, none of the participants had the opportunity to discuss the study with other personnel who had participated earlier, so they had no prior knowledge of the exception items that would be used.

Participants were 91 firefighter recruits from the NSW Fire Brigades in their initial 16 weeks of training. They were all male, between 19 and 44 years old (with a mean age of 30.8 years; SD = 5.2), and had on average 2.4 years’ fire fighting experience (SD = 3.8; range = 0 to 17 years).

The procedure was the same as in Experiment 2, except that (a) there was only one training and test phase and (b) the response options for Question 3 ("What initial attack would you employ?") were modified. The first option was altered to include the possibility of performing a search and rescue. The four options were (a) "Send in crews with hoses for search and rescue and/or to fight fire"; (b) "Send in crews with hoses to protect internal exposures"; (c) "Position crews outside with hoses to fight fire"; and (d) "Position crews outside with hoses to protect external exposures."

The test data were analyzed using a series of repeated measures ANOVAs. Planned contrasts were run, assessing the effects of instruction and practice on the routine items, near exceptions, and far exceptions. The dependent variables were classification accuracy, classification reaction time, and confidence.

Results

Table 5 shows the means, standard errors, and correlations among the variables in Experiment 3. As in Experiment 1, the reaction time measures were correlated with each other, as were the confidence measures. Performance on routine items was positively correlated with performance on far exceptions, suggesting that individuals who learned the rule well were better at recognizing novel exceptions to that rule. Near exception accuracy was negatively correlated with near exception reaction time, suggesting that individuals who recognized familiar exceptions responded more quickly than individuals who did not.

As would be expected, participants were less confident, and took longer to respond, on the far exceptions than on the near exceptions: confidence ($M_{near} = 4.01, SE_{near} = 0.07; M_{far} = 3.83, SE_{far} = 0.07), F(1, 87) = 7.89, p < .05, η² = .08; reaction time ($M_{near} = 5.14 s, SE_{near} = 0.25; M_{far} = 5.80 s, SE_{far} = 0.31), F(1, 87) = 4.58, p < .05, η² = .05. Figure 3 shows the accuracy of participants’ responses for the near and far exceptions in the four groups. Overall, there was no difference between the near and far exceptions in terms of accuracy. Contrary to expectations, the main effects of instruction and practice were not significant. Hypotheses 1 and 2 were, therefore, not supported. However, there was a significant interaction between instruction and exception type, $F(1, 87) = 9.83, p < .05, η² = .10. The exceptions instruction group was more accurate than the routine instruction group on the near exceptions, whereas the exceptions instruction group was less accurate than the routine instruction group on the far exceptions. There were no differences among the groups in performance on the routine items.

Table 5
Means, Standard Errors, and Correlations Among Variables in Experiment 3

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SE</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Routine accuracy</td>
<td>0.75</td>
<td>0.01</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Near exception accuracy</td>
<td>0.81</td>
<td>0.02</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Far exception accuracy</td>
<td>0.78</td>
<td>0.22</td>
<td>.30*</td>
<td>.05</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Routine reaction time</td>
<td>5.88</td>
<td>0.20</td>
<td>—.15</td>
<td>—.11</td>
<td>—.06</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>5. Near exception reaction time</td>
<td>5.14</td>
<td>0.24</td>
<td>—.17</td>
<td>—.27*</td>
<td>—.01</td>
<td>.45*</td>
<td></td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>6. Far exception reaction time</td>
<td>5.81</td>
<td>0.31</td>
<td>—.18</td>
<td>—.11</td>
<td>.47*</td>
<td>.40*</td>
<td></td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>7. Routine confidence</td>
<td>3.97</td>
<td>0.58</td>
<td>.12</td>
<td>—.12</td>
<td>.09</td>
<td>—.10</td>
<td>—.03</td>
<td>—.06</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>8. Near exception confidence</td>
<td>4.01</td>
<td>0.68</td>
<td>.13</td>
<td>—.04</td>
<td>.01</td>
<td>—.06</td>
<td>—.16</td>
<td>—.02</td>
<td>.65*</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>9. Far exception confidence</td>
<td>3.83</td>
<td>0.70</td>
<td>.10</td>
<td>—.04</td>
<td>.02</td>
<td>—.07</td>
<td>.02</td>
<td>—.08</td>
<td>.74*</td>
<td>.55*</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. N = 91.

* p < .05.
Given the unexpected findings regarding performance on the far exceptions, we carried out a post hoc analysis on the far exceptions of the group that received the exceptions instruction followed by routine practice with the group that received the routine instruction followed by exceptions practice. This analysis provides the most direct test of the efficacy of the two training strategies. As can be seen in Figure 3, the group that received the routine instruction with exceptions practice appears to have performed better on the far exceptions than the group that received the exceptions instruction with routine practice. This difference only approaches significance, although it does represent a medium effect size, $r(45) = 1.810, p = .077, d = 0.51$.

**Discussion**

Experiment 3 examined the effectiveness of two alternative strategies for enhancing adaptability: one based on general instruction and the other based on specific practice. The hypothesis that providing general instruction regarding exceptions to a rule will enhance performance on both near and far exceptions was not supported. The exceptions instruction group did perform relatively well on the near exceptions. These items represented exceptions to the rule because there was a danger of imminent collapse or a need for search and rescue. This group had been given information regarding these conditions in the lecture and were able to subsequently recognize cases in which these conditions were present. However, it appears that providing general instruction with respect to the complexity of the domain and the kinds of conditions in which the general rule does not apply caused trainees to become fixated on the specific conditions that were mentioned in the lecture. The exceptions instruction group performed relatively poorly on far exceptions.

To perform well on the far exceptions, the trainees needed to (a) notice the existence of unusual conditions, such as the fact that the building is derelict and due for demolition, and (b) exercise their discretion, by not following the rule that they had learned. The ability to exercise discretion and to modify a rule to suit the requirements of the situation is an important aspect of adaptability that has not been studied previously. These findings suggest that providing explicit instruction with respect to the types of exceptional conditions that might be encountered at test can, paradoxically, impair this type of adaptability.

The comparison between the group that received the exceptions instruction with routine practice and the group that received the routine instruction with exceptions practice is interesting in this respect. The results suggest that presenting exceptions in the context of a set of practice examples might be more effective than explicitly instructing trainees about these exceptions. One explanation for this finding could be that the exceptions may have come as a surprise following the routine lecture and stimulated elaborative and metacognitive processing (Ivancic & Hesketh, 2000). However, care needs to be taken in interpreting this finding, as the trend did not reach conventional levels of significance.

One potential limitation of Experiment 3 is that the exceptions lecture was longer than the routine lecture and may have induced boredom or fatigue. However, we do not believe that this argument can explain our findings or compromise the validity of our conclusions. First, the extra material only added one and a quarter minutes to the lecture, making it unlikely that it would contribute much to boredom or fatigue beyond that already experienced by the participants. Second, even if there were differences in boredom or fatigue, this could not explain why the exceptions instruction group performed better than the routine instruction group on the near exceptions but worse on the far exceptions. Boredom or fatigue would have produced a general decrement in performance.

**General Discussion**

The findings reported in the current article have both theoretical and practical significance. Theoretically, the findings contribute to our understanding of the learning process during critical incident training. Currently, there is substantial interest in the use of critical incident training as a technique for training in decision skills and enhancing situation awareness (Cohen et al., 2000; Klein, 1997). This literature has not been informed by an explicit theory of the learning process. We believe that the same holds true for the personnel training literature in general. By treating critical incident training as a category learning problem, we were able to develop a theory of the learning process and use it to generate a set of testable hypotheses regarding the effects of different design features. We believe that this is where the major theoretical contribution of the article lies. The methodological and practical implications of the studies flow from the application of this theory to the analysis, design, and evaluation of training. In the section below, we explain these implications in more detail.

**Theoretical Implications**

The findings contribute to our understanding of the learning process in two ways. First, the findings suggest that trainees use both rule and exemplar strategies when learning decision-making skills. By itself, this finding is not new, as it has been demonstrated in laboratory settings previously. What is new is the finding that this result generalizes to a training context. Researchers have questioned the relevance and applicability of cognitive research for...
training programs, arguing that the gap between the laboratory and the field is too large for cognitive theory to be applied successfully (Latham & Sejts, 1997). To bridge this gap, we first need to know whether the findings obtained using laboratory tasks generalize to real training contexts. Our results show that they do, suggesting that it is possible to use theoretical accounts of human category learning to describe the learning process during training.

Second, the results supported the prediction that the two learning processes would respond differentially to two training design variables: practice and context. The particular design features that we selected (practice and context) were chosen because (a) they provide a strong test of the predictions of our model and (b) they are important features for trainers to consider. The double dissociation that we obtained in Experiment 2 provides relatively strong evidence to support the existence of two separate learning processes. Dual-process models of the type that are proposed here are always open to criticism on the grounds of parsimony. In the absence of a double dissociation, it is possible that someone may be able to account for the observed effects using a single-process model. This is why it was important to demonstrate differential effects of the hypothetical processes. However, as noted above, the effects of practice and context are also of interest in their own right. Trainers need to make decisions about the amount of practice that they give and the extent to which they attempt to contextualize this practice. It is often not possible to provide as much practice as one would like or to ensure a close match between the contextual features of the training and transfer environments. The current results may help trainers to understand the consequences of these tradeoffs. Of course, further tests of the theory are warranted in which other aspects of the design process are considered. These include factors such as the sequencing of examples, the timing of feedback, and the role of errors as learning events.

Ideally, the role of a learning theory in the instructional design process should be to help the trainer make design choices (Glaser, 2001; Hunt, 2001). Progress toward the development of such theories has been slow. Substantial progress has been made in the description and analysis of the end state of the learning process (experts) and in the development of tools to assess whether an individual has reached this state (e.g., tools for assessing knowledge structures). However, less is known about the learning process responsible for these changes and the way that it responds to design choices made by trainers (Glaser, 2001). Where progress has been made, the focus has been on one specific aspect of the learning process, namely, the transition from controlled to automatic processing. For example, researchers have used theoretical accounts of the transition from controlled to automatic processing as the basis for training needs analysis and training design (Ford & Kraiger, 1995; Rodgers, Maurer, Salas, & Fisk, 1997). Although automaticity is a desirable characteristic under some circumstances, it is not always appropriate or achievable. The theory developed in the current article provides a framework for evaluating a different aspect of the learning process, namely, the types of knowledge that develop during training and the way that those sources of knowledge respond to different aspects of training design.

Methodological and Practical Contributions

The current article makes a methodological contribution by showing how the process dissociation procedure can be used in a training context to evaluate how trainees are learning. The procedure is suited to critical incident or case-based training programs, in which the attributes of the incidents or cases can be manipulated in order to place the hypothetical learning processes in opposition to each other. Potential applications for the procedure include training needs analysis, as well as the design and evaluation of training. By studying how trainees use different learning processes, one may be able to identify training needs more accurately and design training programs that more precisely target these learning mechanisms. For instance, a training program that targets the exemplar learning mechanism might focus on presenting a small number of highly contextualized examples and encouraging trainees to compare and contrast individual examples and to reason by analogy. In contrast, a training program that targets the rule-learning mechanism might present a large number of relatively decontextualized examples and encourage trainees to evaluate each example against a standard set of criteria.

The current article, therefore, addresses the need identified by Ford and Kraiger (1995) for incorporating new cognitive constructs and techniques into the training literature. It is the focus on the learning process itself that differentiates the process dissociation from other cognitively oriented techniques that have been introduced to the training literature recently.

It should be noted that there is at least one other approach for evaluating learning processes within the cognition literature that, in principle, could also be introduced to the personnel training literature. Computational modeling is commonly used to draw inferences about underlying cognitive processes. John Anderson and his colleagues have been using a cognitive modeling architecture called Adaptive Control of Thought–Rational (Anderson & Lebiere, 1998) as the basis for designing training programs for skills such as LISP programming and high school mathematics (Anderson, Douglass, & Qin, 2004). Although this research program is undoubtedly valuable, the modeling is very resource intensive and requires advanced mathematical skills. We believe that the process dissociation procedure is simpler and easier to apply in personnel training contexts.

From a practical perspective, the current findings illustrate an important dilemma for trainers. On the one hand, one of the goals of training is to encourage generalization. Our findings suggest that trainees will use both rules and examples to enable transfer to new cases. On the other hand, trainers wish to avoid trainees generalizing a rule or an example to cases in which it does not apply. Our study demonstrates that trainees can generalize both rules and examples inappropriately. Experiments 1 and 2 demonstrated that examples can not only facilitate performance but also impair performance. The trainees were prepared to generalize from an example even if that example belonged in the wrong category. These effects were quite strong when the contextual features of the training and test examples were closely matched and were not counteracted by simply providing more practice. This represents a dilemma for trainers, because on the one hand, they may be able to enhance generalization by increasing the contextual overlap between the training and transfer environments, but on the other
hand, this increases the risk that trainees may generalize inappropriately.

Experiment 3 suggests that instructing trainees about the conditions under which a rule does not apply may be counterproductive. The results suggest that this strategy either impaired trainees’ ability to recognize exceptions that they had never seen before or reduced their tendency to exercise discretion in the application of the rule. If trainers need to present information about exceptional conditions, then it may be better to do so using specific incidents rather than via general instruction.

Strengths and Limitations

There are a number of strengths of the current series of experiments. A major strength was the use of experimental designs that provided a high degree of control and allowed us to draw inferences regarding the decision processes that trainees were using. Furthermore, the process dissociation technique provides a practical tool that trainers can use to assess how their trainees are learning. A second strength was the use of trainee firefighters as participants, rather than university students. The trainees had, on average, 2–3 years’ experience. The studies addressed issues that were of immediate concern to them. The use of a personnel training context provided an opportunity to obtain tight experimental control, but within an environment where trainees were motivated and the training and test items were highly relevant.

On the other hand, there are at least five potential limitations that need to be considered. First, it is not known whether the training strategies examined in the current article had any lasting effects on job performance. However, Experiments 1 and 2 involved within-subject designs, so we would not expect to observe differences in subsequent job performance. Experiment 3 did have a between-subjects design; however, for ethical reasons we debriefed all trainees at the end of the study in order to minimize any differences in subsequent job performance attributable to the experimental manipulation. It would be inappropriate to allow recruits into the field if they had been given training that was considered, on a priori grounds, to be suboptimal. Furthermore, it is important to note that job performance is not the only criterion by which training strategies should be evaluated. There is a wide range of environmental factors that attenuate the effects of any training intervention over time. The focus of the current set of experiments was on the way in which people learn during training and their ability to adapt to problems that they have never seen before. These questions are of practical significance to trainers regardless of whether the work environment supports transfer.

Second, the current experiments were limited in the amount of information that could be provided in the lecture, the opportunity that participants had to study this material, and the amount of practice that they received. For example, the lecture lasted only 15 min, and there were only eight practice examples. For this reason, the current findings apply only to the early stages of skill acquisition. However, it is important to note that NSW Fire Brigades recruits are currently not given any training of this form. The information included in the lecture, and the types of practice examples that we used, are currently used for training at higher levels in the fire service (team leaders and incident commanders). Therefore, the training program that we developed is actually more complex and more detailed than the training that recruits currently receive. We believe that the length and complexity of the training program that we developed is appropriate, given the needs of the trainees and the context in which the training occurred.

Third, the task that we used in this study required entry-level firefighters to make tactical decisions. In the field, team leaders and incident commanders make tactical decisions, not entry-level firefighters. For example, the trainees were asked to indicate what the exposures were (i.e., where the fire was likely to spread) and where they would attempt to contain the fire. However, the key decision that we were interested in was whether the firefighters would enter the building. Firefighters do need to be able to assess whether a building is safe to enter. For this reason, the fire service that we were collaborating with considered it important for their recruits to be provided with early exposure to the issues.

Fourth, the training program that we used in these studies presented simulated scenarios, which do not involve the same level of risk and stress as would be encountered under real conditions. To carry out the studies in real fire situations would involve sacrificing the experimental control essential for the application of the process dissociation procedure.

The final potential limitation concerns the materials that were used. It is possible that the effects observed in the current experiments reflect the particular way in which the lectures and examples were constructed. For example, it is possible that the examples may have been seen as artificial, particularly with respect to the contextual features. However, the examples were constructed by subject matter experts and are similar to the kinds of examples used in other training programs in the fire service.

Conclusion

The three studies demonstrate how the process dissociation procedure can be used to evaluate how trainees learn decision-making skills in a critical incident training program. This procedure is grounded in an explicit theory of the processes that are responsible for learning. The three studies have shown that trainees will use both rules and examples to help them generalize to new problems, and that under some circumstances they will make errors when they do so. Of course, errors can also be useful as learning events in their own right. Studies are currently underway examining how to use error examples in developing adaptive decision making in fire control (Joung, Hesketh, & Neal, 2006). More generally, we hope that an explicit theory of the learning process will be useful for trainers who wish to apply cognitive constructs and principles to the analysis, design, and evaluation of personnel training programs. In this way, it may be possible to bridge the gap between basic research and the concerns of researchers and practitioners working in the field.

References

Appendix

Structure of Training and Test Items Used in Experiments 1 and 2

<table>
<thead>
<tr>
<th>Training items</th>
<th>Test items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building structure Fire size Intensity of fire</td>
<td>Building structure Fire size Intensity of fire</td>
</tr>
<tr>
<td><strong>Matching items</strong></td>
<td></td>
</tr>
<tr>
<td>1a Resistant Small Low</td>
<td>1b Resistant Small High</td>
</tr>
<tr>
<td>2a Not resistant Small Low</td>
<td>2b Not resistant Large Low</td>
</tr>
<tr>
<td>3a Resistant Large Low</td>
<td>3b Not resistant Small Low</td>
</tr>
<tr>
<td>4a Resistant Small High</td>
<td>4b Not resistant Small High</td>
</tr>
<tr>
<td>5a Resistant Large High</td>
<td>5b Resistant Large Low</td>
</tr>
<tr>
<td>6a Not resistant Large Low</td>
<td>6b Not resistant Large High</td>
</tr>
<tr>
<td>7a Not resistant Small High</td>
<td>7b Not resistant Small Low</td>
</tr>
<tr>
<td>8a Not resistant Large High</td>
<td>8b Resistant Large High</td>
</tr>
<tr>
<td><strong>Novel items</strong></td>
<td></td>
</tr>
<tr>
<td>9 Resistant Large High</td>
<td></td>
</tr>
<tr>
<td>10 Resistant Large Low</td>
<td></td>
</tr>
<tr>
<td>11 Not resistant Large Low</td>
<td></td>
</tr>
<tr>
<td>12 Resistant Small High</td>
<td></td>
</tr>
</tbody>
</table>

Received June 23, 2004
Revision received August 29, 2005
Accepted October 26, 2005


