Combinatorial semantics in the visual world: A representational account of real-time event processing

Christopher Charles Rowson

PhD

Psychology

February 2011

Abstract

The experiments presented in this thesis sought to establish combinatorial effects operating within novel semantic priming paradigms. These experiments demonstrate the combining of several pieces of semantic information (specifically, what certain items afford in terms of some form of interaction) from objects within simple visual scenes, and an account is offered based on the construction of representations from these presented items. These composite representations subsequently facilitate responses to target items, but it is only with a combination of items (e.g. a knife and a tyre) that is felicitous to a particular event (e.g. bursting) that priming to target items (e.g. a burst tyre) obtains. Priming is less evident when a combination of items is presented that does not allow the same event to occur (e.g. a ruler and a tyre). These combinatorial processes appear to occur in longer time windows than those of automatic visual processes such as gist abstraction, and the representations activated upon viewing these stimuli appear to be specific to the objects presented.

List of tables and figures

Acknowledgements

I would like to thank my supervisor Gerry Altmann for all his help and support throughout the duration of the degree course. I would also like to thank Silvia Gennari, and the members of my research committee – Gareth Gaskell and Philip Quinlan for their valued input.

Declaration

This thesis contains original work completed solely by the author under the supervision of Prof. Gerry Altmann.

Chapter 1

The notion of internal (mental) and external (physical) worlds and how entities around us are coded or internalised mentally has been a major concern of cognitive psychology since its inception. The methodologies used to explore this have provided researchers with ways of studying the complex interplay between vision, language and action. The current chapter aims to provide an overview of literature relevant to the studies that are to be presented and discussed in subsequent chapters. Each chapter will be organised in the following way: First, the issue that the experiments aim to address will be considered with a brief discussion of the relevant supporting studies. Next, the methods and information relating to the design of the experiment will be provided. Finally, data from the experiment will be presented with a discussion of the relevant implications and edifications that these data provide.

This thesis will explore several questions related to the aforementioned notion of internal and external worlds, and how information pertaining to both is processed and shared between them: How is the external world represented cognitively, and what form do these representations take? Is information relating to features of objects or entities routinely combined so as to aid subsequent interaction with them? And what are the approximate time windows in which these representations are constructed and utilised?

To begin with, vision research has perhaps offered the earliest insights into the nature of the representations activated as our eyes scope our visual environment. It has also provided indications as to the types of processes involved with this extraction of information from our surroundings, and what type of information is extracted in the first place.

Vision and scene gist

Our perceptual experience as we look around our visual environment is one of a stable, richly-detailed world, with everything in our visual field experienced simultaneously. Intuitively, it would seem we build up an equally detailed inner ‘picture’ of that environment, a mental representation with which to guide appropriate visual operations (Rensink, 2000). Indeed, many early attempts to specify the role of the visual system converged on the idea that vision’s primary function is to recreate a detailed mental copy of the external world (Tsotsos, 1987). However, there is a considerable body of evidence that contends this conceptualisation of vision as some sort of ‘scene recovery’ system (Aloimonos & Rosenfeld, 1991), suggesting instead that we never build up a complete representation of everything in our visual field (e.g. Milner & Goodale, 1995; Rensink, 2000; Rensink et al., 1997).

According to Clark (1999), the main function of the visual system is to use visual information ‘cheaply’, only utilising information if it is needed *at the time*, to aid our real-time interaction with the external world. Earlier views of the visual system, which present it as a means by which one can ‘throw the world away’ (Clark, 1999, p345), are thus rejected by researchers aiming to reconceptualise the process of extracting visual information from our environment. Researchers in the field of interactive vision present an alternative view of this system, whereby the active retrieval of information is driven by a dynamic, changing concurrent visual scene.

Even though it appears we do not build up detailed representations of our surroundings, we can still derive semantic information from visual stimuli in the form of ‘gist’, within a few hundred milliseconds. The simple fact that we don’t build detailed representations might be one of the reasons we *are* able to extrapolate semantic information rapidly from visual stimuli – The system is not laden with the processing of unneeded information.

Although it is a fairly poorly-defined term in the literature, Henderson and Ferreira (2004) outline three elements involved with the abstraction of gist from visual scenes. The ‘identity’ of a scene is established, along with semantic information associated with this scene and scenes of a similar nature, and also information serving to establish the global spatial layout of the scene. Henderson and Ferreira suggest that gist can be thought of as the ‘general semantic interpretation’ of a visual scene.

Perhaps the most striking thing about gist is the speed with which we appear to extract it from visual stimuli. Various studies have suggested that gist can be picked up in as little time as 100 ms (Beeckmans, 2000; Biederman, 1981; Oliva, 2005; Potter, 1976; Rensink, 2000). Potter (1976) for example, found that even when the presentation times of singly presented pictures was short, at less than 120 ms, participants were still able to accurately detect them in a later recognition task.

The technique used by Potter to explore scene gist became known as Rapid Serial Visual Presentation (RSVP) (Potter & Levy, 1969; Potter, 1975, 1976; Potter, 1999), in which a sequence of scenes (usually colour photographs) are presented at rates roughly similar to eye fixations. In Potter's (1976) paradigm, the participants' task was either to recall the pictures in a memory test, or to search for a target picture in a sequence. Potter's results demonstrated that even when targets had been indicated by verbal labels (instead of pictures) in the search condition, participants' detection performance was far better than in the memory condition, even at fast presentation rates of less than 120 ms. Potter suggested that the gist of a scene could be abstracted within this time, but that the memory trace for these stimuli decayed rapidly. With regard to the accurate detection of targets that were only specified with a verbal label, Potter concluded that detection in the search condition was not dependent upon participants using the visual information associated with the target to accurately pick it out of a sequence.

Biederman's (1981) work concerning gist abstraction also demonstrated participants' ability to extract gist with stimulus presentation durations as little as 100 ms. In a paradigm often used by researchers in the field of visual cognition, Biederman presented participants with a scene which was quickly replaced by a pattern mask. Prior to the scene appearing on the screen, a target label was presented to participants, with a spatial marker cue appearing just after the scene had disappeared. Participants were required to establish whether the target object had appeared in the location defined by the marker. The target label did not appear on every trial, and was not always consistent with the visual scene. However, Biederman found that participants were more accurate if the target label was consistent with the scene, and they were able to derive the gist from scenes at short presentation times of 100-200 ms.

Such findings raise the question of how we are able to abstract the meaning of a visual scene so quickly, without forming adequate representations of everything in it. Rensink (2000) proposed three independent systems for the processing of visual information, of which gist was a part of the first. Before stable representations could be constructed by the second-level system, the ‘volatile’ structures of the concurrent scene are extracted in the form of gist. Rensink suggested that this extraction of the overall meaning of a visual scene could then be used to direct one’s attention to objects of interest within the visual field. Once this has been achieved, the third-level system is able to facilitate the *perception* of objects.

Schyns and Oliva (1994) sought to further refine the time-course of gist abstraction, and found that the different types of gist from a hybrid of two overlaid scenes can be affected by presentation rates. In these studies, low spatial frequency (LSF) and high spatial frequency (HSF) images were combined to form a single ambiguous scene. The LSF image in this case was an image of a street in a busy city, and the HSF image was a hallway. Participants were asked to complete a categorisation task with these hybrid ambiguous images, where they were required to judge whether it matched a subsequent normal image (the target). Schyns and Oliva found that at presentation rates of 30 ms, participants performed matching based on the images’ coarse structures (the city scene), but at longer rates of 150 ms, matching tended to be based on fine structures (the hallway).

Dell’Acqua and Grainger (1999) provided a notable demonstration of the rapidity with which semantic information can be derived from visual stimuli. These authors’ work demonstrates the activation and overlap of representations engendered by visual and linguistic information. In this study, participants were presented with line drawings of ‘artifactual’ objects (utensils, vehicles etc.) or natural kinds (mammals, vegetables etc.), which were identical, related or unrelated to a subsequently presented target word. Participants were required to make a judgement as fast as they could as to whether the word refers to something that is an artifactual object or a natural kind. These authors found that even with picture exposure durations as short as 17 ms, participants were able to categorise objects much faster when the objects were either identical or related, compared to when they were different from the targets. These findings suggest that unconsciously presented objects can activate semantic category information, even though they have received scant attention.

Dell’Aqua and Grainger’s results indicate that word recognition can be facilitated or ‘primed’ with the prior presentation of a visual stimulus, even at presentation times so short, that participants report not having seen the them. Oliva & Torralba (2007) argue that an important facilitator of object recognition is context. If a target has been preceded by something which seats it within a contextual frame of reference, then responses to that target are primed. For example, participants' categorisation responses to a target (e.g. a loaf of bread) are facilitated if it has been preceded by an appropriate context (e.g. a kitchen counter). This is known as the scene consistency-inconsistency effect (Palmer, 1975).

In a recent finding supporting this notion of contextual priming, Green and Hummel (2006) discovered that the relative positions of objects to one another influences object perception. In their first experiment, participants were asked to respond as to whether the second picture in a two-picture sequence (both presented for just 50 ms) matched a label given prior to that. For example, they might have been presented with the label ‘glass’, then a picture of a jug positioned as if it were to pour liquid into a glass, and then a picture of a glass on which they were to make their decision. The objects were related (as in the jug and glass case), unrelated (e.g. a key and a glass), interacting (with the jug positioned so as to pour liquid into the glass), not interacting (objects positioned with no interaction), positive (the second object matched the label) or negative (the second object did not match the label).

Green and Hummel found that participants were significantly faster to respond when the jug interacted with the glass than when it did not. However, this performance disappeared when the ‘interacting’ objects were unrelated. In order to rule out any postperceptual account for their above data, Green & Hummel increased the stimulus onset asynchrony (SOA, the time between the onset of the prime and the offset of the target) to 250 ms. These authors suggested that if the above results are due to postperceptual processing, then they should persist with longer SOAs. If, as the authors maintain, the effects are due to perceptual grouping, then these effects should be eliminated or significantly reduced at a longer SOA of 250 ms. As predicted, significantly less priming was evident for these stimuli at longer SOAs.

These authors’ third experiment was designed to counter the argument that the effects arise from attentional cueing: It may be that the position of the jug forced participants’ attention to the position on the screen where they expected the second object to be. In this experiment the two objects were presented in the opposite order – so participants received the glass first, then the jug. The results reported for this third experiment point to perceptual grouping rather than attentional cueing because if it was the latter, one would expect a striking difference between the results of experiments 1 and 3. As this was not the case, Green and Hummel were able to argue that functionally interacting objects are perceptually grouped and that, consonant with previous research, object recognition is at least in part driven by the context in which objects are presented.

Semantic priming

The role of context in studies of linguistic processing is perhaps even clearer, and the benefits provided by context for the interpretation of linguistic expressions is well documented (e.g. Tabossi, 1988; Williams & Colombo, 1995; Smith et al., 1994). For example, Tabossi (1988) demonstrated priming effects for combinations of words where context in the sentences played an important part. Tabossi found that the target word ‘thorns’ was primed by sentences such as ‘the girl was pricked by a rose’, but not sentences such as ‘the girl smelled a rose’ or ‘the girl was pricked by a wasp’. Here, the context of the sentence seemed to have a clear impact of facilitation on responses to related target words. In addition, Tabossi found that participants responded faster to the target word ‘fat’, when presented with an appropriate priming sentence such as ‘In order to follow her diet, the woman eliminated the use of butter’, than to either a neutral sentence: ‘Before paying, the man checked the price of butter’, or an inappropriate priming sentence: ‘In order to soften it, the woman heated the piece of butter’. Indeed, these findings have been supported by a number of studies (e.g. Andrews et al.,1993; Seidenberg et al., 1982) and crucially, there is also evidence to suggest that these priming effects are not due to any one word in the sentence (e.g. Potter & Faulconer, 1979).

The combining of linguistically presented concepts has also been explored within a framework of conceptual combination and compositionality (Tabossi et al., 2008; Kamp & Partee, 1995). Language is compositional in that the formation of sentence structure is done so with a particular ordering of its constituent elements. Within this framework, the meaning of sentences is gleaned from, firstly, the individual meaning of the words that make up the sentence, secondly, the syntactic arrangement of the sentence, and thirdly, the context in which the sentence is provided to the listener or reader.

Conceptual combination describes the ways in which one's arrays of available concepts are combined to construct higher-order concepts and develop complex representations that encompass them. Estes and Glucksberg (1998) for example, found that semantic priming could account for the effect of context on the accessibility of the features of combined concepts. These authors provided participants with context passages, before asking them to make true/false judgments on subsequent probe and comprehension sentences. These authors found that participants were faster to make their responses when particular features of the sentences (requiring combination) were primed by the preceding context passages.

As robust priming effects have been established from tasks of both object and word recognition, it is not surprising that the notion of a single store of semantic information for both visual and linguistic codes has been proposed (Chase and Clark, 1972). Vanderwart (1984) found that pictures prime words to the same extent as words themselves, which provides support for this single-code view of semantic access. In Vanderwart's paradigm, participants were presented with a word or a picture which preceded the presentation of a target word. The prime was either (i) an identical concept (e.g. apple-apple); (ii) semantically related (e.g. apple-pie) or (iii) unrelated to the target words (e.g. apple-lamp). All stimuli were presented visually, and participants were required to make a lexical decision on the targets. The single-code view would predict that, where identical concepts have been used, reaction times to the primes should be roughly the same. Where semantically-related concepts have been used however, the single-code view would predict less facilitation than for identical concepts because in the latter case, the abstract code for the same concept will be activated by both the prime and the target.

In fact, Vanderwart found this to be the case. Participants' lexical decisions were the fastest for repetitions, the next fastest for related concepts, and the slowest for unrelated concepts. Crucially, this progressive facilitation as the prime became more related to the target was almost identical for both pictures *and* words, indicating that pictures prime words to the same extent as words themselves. These results suggest that not only do vision and language interact in both directions, but linguistic and visual information access a single semantic system, consonant with a single-code view of access to semantic knowledge. Later studies (e.g. Bajo & Canas, 1989) have also provided support for the activation of common semantic and phonetic representations for pictures and words.

Although priming paradigms reveal something about the types of semantic information activated as a result of related linguistic or visual input, their impact has been diminished by a recurrent problem. Koriat (1981) was the first to suggest that these effects may not be due to facilitation from the prime at all – It may be that the contextual 'appropriateness' of a prime could be retrieved *after* participants have been presented with the target. Koriat argued that the target 'reactivates' the prime, and it is only until both stimuli have been processed that a relation between them is established. Koriat referred to this facilitation from the target to the prime as backward priming. According to association norms, 'baby' is associated to 'stork', but 'stork' is not associated to 'baby'. Koriat found that, when asked to perform a lexical decision on the target for forward associations such as stork-baby, as expected, participants' reaction times in a lexical decision task (LDT) were facilitated compared to controls. However, the same degree of priming was found for backward associations such as baby-stork, which is problematic for the traditional accounts of semantic priming.

It is still not clear, however, whether these problems of backward priming are due to something intrinsic about priming itself, or something intrinsic about the task. Peterson and Simpson (1989) presented observers with a prime, and then a target on which they had to make either a lexical decision or complete a naming task. Words were either unidirectional (Stork-Baby) or bidirectional (Baby-Cry) associates, and were presented in either forward or backward orders. In Peterson and Simpson's first experiment, participants made a naming response to the target, and in their second experiment, participants made a lexical decision. Forward priming effects were reported in both experiments, but only in the second experiment, where the lexical decision had been used, were there effects of *backward* priming. These results confirm Seidenberg, Waters, Sanders and

Langer's (1984) finding that backward priming effects arise with the LDT, but not with naming tasks.

Embodied cognition and affordances

Priming effects are not limited to paradigms in which responses are made on the basis of previously seen stimuli. Recent evidence has demonstrated priming not to words or pictures, but to actions. If one recalls from previous discussion that we do not seem to build up detailed 3-D representations of our surroundings, then the fact that motor responses are facilitated as we process visual stimuli in priming paradigms seems consonant with such a view.

Embodied cognition seeks to redefine the nature of cognition, shifting the perspective away from a focus on abstract symbols onto 'situated activity' (Anderson, 2003), where processing is seen as part of one's active participation in a rich, complex visual world. More specifically, the world around us is computed in terms of 'patterns of possible interactions' (Glenberg & Robertson, 2000) that serve to aid the goal-directed behaviour of situated entities.

Much of the literature concerning embodied or 'grounded' cognition focuses on establishing the interplay of language, perception and action within a situated perspective. For example, objects' orientations may have an influence on how they are processed due to the configuration of our own bodies (facing away from the object, right-handed or left-handed, and so on). Some researchers even go so far as to suggest what an object *means* to us is what that object allows us in terms of some kind of interaction (e.g. Glenberg, 1997).

In a study that perfectly illustrates this consideration of processing as conducted by active entity, Craighero, Fadiga, Umiltà & Rizzolatti (1996) found that participants' performance on a grasping task was facilitated by a previously seen rectangle whose orientation matched that of the object to be grasped. Some of the most notable finds in the literature include those where the demands of the task do not even require the motor action to be carried out. For example, there is a large body of evidence which demonstrates the link between the activation of motor neurons involved with controlling tool use in monkeys and the activation of those same neurons when the same tool use is merely *observed* (Grafton, Fadiga, Arbib, & Rizzolatti, 1997; Murata, Fadiga, Fogassi, Gallese, Raso, & Rizzolatti, 1997). Similarly, Tucker and Ellis (1998) reported that participants' response times were facilitated if the response was made with the same hand as that which would be used to grasp a previously seen object (e.g. a teapot).

Finally, Glenberg and Kaschak (2002) found that participants are faster to respond in a sensibility judgement task to sentences that describe an action either towards, or away from the body if the response requires them to make an arm movement in the same direction. For example, if they are required to make the judgement on the sentence 'Open the drawer', then they are quicker to respond if they need to make an arm movement *towards* their body. Participants were faster to respond to converse sentences such as 'Close the drawer', but only when the response requires them to make an arm movement *away* from the body. This is known as the 'Action-Sentence Compatibility Effect' (ACE).

Closely tied with this notion of action priming is the argument that the activation of event knowledge is crucial to our understanding of the world around us (Glenberg, 1997; Barsalou, 2008). Indeed, some authors claim that we process visual information in terms of a pattern of potential events or interactions with our surroundings. According to Glenberg (1997) what an object, event or sentence *means* to us is what we can do with it in terms of some kind of interaction.

A central construct in Glenberg's 'situated' approach to cognition is that of the affordance. Affordances are simply the types of interactions that an object or entity allows because of particular surfaces or features inherent in them (see Gibson, 1979). While these affordances reflect the different ways an object can be used, they also change in accordance with the context in which one uses it. For example, a glass affords being filled with liquid, but this affordance changes if the glass is left upturned.

Glenberg describes the notion of object affordances in relation to the theory of 'mesh'. As our surroundings could be computed in terms of patterns of possible actions, these patterns can be integrated or 'meshed' together and brought to bear on our comprehension of the visual environment. These meshed action patterns are constrained by the types of things an object or entity affords with regard to a particular event. These patterns are also described as being very dynamic, and readily change in accordance with changing action patterns.

This notion of meshing knowledge and representations is central to the work presented in this thesis. The combining of affordances forms part of the motivation for each current experiment: Do we readily combine semantic knowledge from two unrelated objects to form a composite representation of the event that, together, these items allow? This question is to be explored in detail in the work that follows.

Event knowledge and the visual world

In the current work, the combining of semantic information and object affordances is assumed to give rise to the activation of relevant event representations. These event representations can be used to aid subsequent processing of stimuli that is associated with the activated event. Hommel, Müsseler, Aschersleben and Prinz (2001) installed the notion of event representation into their theory, encompassing representational domains, codes and systems. Within the 'Theory of Event Coding' (TEC), representations of events are assumed to be contained within a single representational domain, with information pertaining to any 'to be-perceived' or 'to-be-generated' event. These representations not only support perceptual systems, memory, and reasoning, but also action-based functions (such as action planning, control, and conduction). As a fundamental tenet of this theory is that perception and action representations are not stored separately, perceptual codes and action codes are said to prime one another because of their overlap in this single representational domain. Hommel et al. describe perception and action planning as 'intimately related' and able to give rise to an unimaginable number of interactions within our environment.

Implications concerning the organisation of these types of event representations in memory were provided by McRae, Hare, Ferretti and Elman (2001). This research investigated the activation of verbs from nouns by way of event representations, and made use of typical agents, patients, instruments, and locations as primes. By 'typical', these authors mean the types of agents, patients and so on, that would typically be used for a particular event. For example, a typical agent for the event ‘praying’ might be ‘nun’. These authors predicted that the nouns used would prime verbs associated with the event typically used in conjunction with whatever the noun refers to. McRae et al. gave participants nouns which they were to read to themselves silently. They were then asked to say the verb that appeared next as quickly and accurately as they could.

As predicted, when naming verbs after typical agents (e.g. nun, praying), patients (e.g. dice, rolled), instruments (e.g. shovel, digging), and locations (e.g. arena, skating), participants were significantly faster in their naming responses, compared to when nouns which were not typical of their respective event were used. The results suggest that event memory is organised in such a way that when a particular object (or a word denoting an object) is seen, information corresponding to the types of events that it is typically used for, or in conjunction with, is activated.

Another effective method for investigating the activation of event knowledge has been the use of tasks involving the ‘Visual World Paradigm’. The basic setup for this paradigm is to track participants’ eye movements over a concurrent visual scene as they hear a sentence referring to constituent elements of that scene. For example, Altmann and Kamide (1999) used a scene consisting of a boy, a cake, and various other distractor items. These authors found that, when participants’ eyes were tracked during the sentence ‘The boy will eat the cake’, more saccades were made to the cake during the word ‘eat’, than any other item. When ‘eat’ was replaced by something more general however, like ‘move’, participants did not show a preference for any particular item during the acoustic lifetime of this replacement word.

This was, the authors argued, because the cake was the only edible item in the scene and participants were showing evidence of anticipatory eye movements. Rather then event knowledge being activated upon seeing a noun that could engage in, or be used for, a particular event (as in McRae et al.'s 2001 experiment) here, participants are anticipating what could be referred to based on the affordances of the items in the scene – The cake is the only item which affords eating, and so this item is anticipated on the basis that the boy will eat something in that scene.

This demonstration of anticipatory eye movements was expanded by Kamide, Altmann, and Haywood (2003). Utilising the same Visual World Paradigm, scenes were presented to participants that contained, for example, a man, a girl, a motorbike, and a fairground carousel. The idea here was that the man would be more likely to ride the motorbike, while the girl would be more likely to ride the carousel. Upon hearing the sentence 'The man will ride the motorbike', more eye movements were made towards the motorbike *before* the word 'motorbike' was actually heard. Similarly, when hearing 'The girl will ride the carousel', more eye movements were made towards the carousel before the acoustic onset of the word 'carousel'. What distinguishes this experiment from the Altmann and Kamide (1999) experiment is that it shows the anticipatory eye movements were not simply a product of the association between the verb 'eat' and cakes, but they were born from the *combination* of that particular verb with its agent (the boy).

Current research

The experiments to be presented and discussed in subsequent chapters will focus on many of the questions and considerations raised in this overview of supporting literature. We know that semantic information activated in sentence comprehension tasks can be combined to aid facilitation of a target, but is this type of operation exclusive to the linguistic domain, or can a similar process operating in the visual domain be established? If so, under what conditions (e.g. exposure durations, sequencing, task demands) would these effects be borne out? There is also the notion of causality, evident in many of possible interactions between objects, that warrants investigation: Do these representations code the changes in state of certain objects resulting from some form of causal-based interaction?

The next chapter will present two experiments that both utilise a cross-modal semantic priming paradigm to establish combinatorial effects operating upon visual stimuli. These effects refer to the combining of semantic information from two unrelated objects, and the resultant facilitation of responses to targets which represent a common event. In these experiments, participants' response times in a lexical decision task will be measured to an auditory target such as the word 'drink'. These targets will be preceded by sequentially presented pictures which, in prime trials, afford the event denoted by the target, such as with 'drink', a bird and a puddle.

Subsequent chapters will expand upon the notion of combinatorial semantics introduced in the following chapter, and these effects are to be established in purely visual paradigms. An attempt will also be made to elicit these effects at shorter stimulus exposure durations, in experiments which provide participants with a causal event structure. Experiments are then to be presented which focus on the form with which these activated representations take.

Chapter 2

The current chapter contains two experiments, both utilising a semantic priming paradigm to examine whether the affordances of visually presented items are routinely combined to form event-based representations, and what the time-course of such a process might be. Firstly, literature relevant to the current set of experiments will be described. Secondly, the methodologies and findings from these experiments will be presented. Lastly, an examination of these findings with respect to their implications for related literature will be provided.

To begin with, we know that combinations of words facilitate responses to a common related concept (Tabossi, 1988). And, as pictures appear to prime words to the same magnitude as words themselves (Vanderwart, 1984), a logical progression would be to investigate whether combinations of objects in a visual scene can facilitate activation of concepts related to the *combination* of items in the scene; i.e. over and above the activation of just individual object concepts.

A study which could be considered the visual equivalent of Tabossi's (1988) experiment was conducted by Altmann, Charles and Gennari (2005). Participants were presented with spoken verbs, each one preceded by one of three corresponding scenes (see figure 2.1). In each trial, participants were required to view one of the scenes (presented for 1750 ms), and then respond to the corresponding target in an auditory lexical decision task. There were three conditions, and although participants received stimuli from each condition, they were only presented with one of the three corresponding scenes for each target word.

Prime No-agent No-theme

Fig. 2.1. *An example of scenes corresponding to the target 'drink' in Altmann, Charles & Gennari's (2005) study.*

The three scenes accompanying, for example, the target word 'drink' are shown in figure 2.1. The 'prime' condition consisted of a scene comprising a bird and a puddle (priming might occur in this case because the combination of the bird and the puddle affords a drinking event); the 'no-agent' condition was a scene of a boy and a puddle ('no-agent' because there was nothing in the scene that was likely to drink from the puddle); and finally, the 'no-theme' condition, comprising the boy and the bird ('no-theme' because there was nothing to drink from). It was predicted that the combination of the bird and the puddle in the prime condition would speed responses to the target 'drink', and there would be less facilitation when the puddle was presented with the boy (the 'no-agent' condition) or when the boy was presented with the bird (the 'no-theme' condition).

The authors' predictions held, and participants were significantly faster to respond when shown the scene with the bird and the puddle, compared to when the bird was shown without the puddle, or the puddle was shown without the bird. In this case, it was only the *combination* of the bird and puddle which activated a representation of a possible drinking event. As in many of the visual world experiments, these findings indicate the overlap of representations engendered by the visual information in the scenes, and representations activated by the linguistic information conveyed by the target verbs.

Although the interaction of vision and language is clear here, the mechanisms which allow such interplay are less apparent. Are these event-based representations activated within exposure durations similar to that of the time course of gist abstraction, and do these priming effects merely arise from strategic, rather than automatic processes, given the long exposure duration of 1750 ms? (see McNamara, 2005). As we know, semantic priming has been observed even in cases where very short exposure durations of 17 ms are used, and participants do not report seeing the prime (Dell'Acqua & Grainger, 1999). Moreover, even when memory for rapidly presented pictures is not stable, conceptual information relating to them is activated (Potter, 1975; 1976).

Experiment 1[1]

In addition to the conditions of 'prime', 'no-agent' and 'no-theme' used in the study conducted by Altmann, Charles and Gennari (2005), the present experiments used an additional baseline condition, in which experimental items were neither related to the target, nor related to each other. This provided an effective comparison for the other conditions, as no priming should be expected in this case at all. Also, in the other conditions, it may have been that priming was observed from just one experimental item on its own. So, in the example of

---

[1] Experiment 1 was conducted as a partial requirement of the degree of MSc in Reading, Language and Cognition. Although data for this experiment were collected before the PhD course had begun, they have been included here for the sake of clarity.

‘drink’, priming may arise from seeing either the bird on its own, or the puddle on its own. In this baseline condition, however, there is nothing to potentially prime the target word. Figure 2.2 shows an example of the stimuli used in the current experiments for the target word ‘drink’.

Prime

No-agent

No-theme

Baseline

Fig. 2.2. *Stimuli presented in each condition for the corresponding target word, ‘drink’.*

The stimuli in the current experiment were presented sequentially to participants (i.e. the bird first, then the puddle) because of a design issue with the Altmann, Charles and Gennari (2005) study. One reason posited by these authors as to why previously reported effects seemed to disappear with shorter stimulus exposure durations was that each object or entity to be processed was not presented in the centre of participants’ visual fields. For example, when a bird and puddle in a single scene was presented to participants for the event ‘drink’, the bird appeared to the right side of the puddle, requiring an additional saccade and fixation in on order to process all the required information in the scene. This may have resulted in a cost for the processing time of the scene.

In the current experiment, it was predicted that the combination of experimental items would facilitate participants' responses to a related target verb in lexical decision. As in the previous study, it is likely that less priming would be observed in the 'no-agent' and 'no-theme' conditions, whereas for the baseline, responses were expected to be slower than those in any of the other conditions. The combining of information is assumed to occur during the presentation of the second picture. Consequently, the duration of this second picture was manipulated (250 ms in Experiment 1; 100 ms in Experiment 2) while keeping the duration of the first display constant (2000 ms).

Method

Participants. 36 participants from the University of York student community took part in the experiment, and were given a sum of £2 for participating. All were native English speakers and had normal or corrected-to-normal vision.

Materials

Visual stimuli. There were four picture pairs for each of the 28 target words, which represented the 'prime', 'no-agent', 'no-theme' and 'baseline' conditions (see appendix 1). For example, the picture pairs for the target 'drink', consisted of a bird and a puddle (the prime), a book and the puddle ('no-agent', there is nothing that will drink from the puddle), the bird and the book ('no-theme', there is nothing for the bird to drink from), and the book and a toothbrush (baseline, objects completely unrelated both to the target, and each other). Each target word had its corresponding ‘prime’, ‘no agent’, ‘no theme’ and ‘baseline’ conditions. In addition, there were 112 filler pairs, each accompanied by either filler nonwords, adjectives or nouns. The items were downloaded from www.clipart.com, and were shown at a resolution of 640×480 on the 17” monitor of an Apple iMac computer.

Each participant saw all filler items but only one of the four corresponding experimental pairs for each target word. Participants were assigned to one of four groups, constructed using a Latin Square. For the ‘drink’ example, therefore, the first participant (assigned to group 1) would see the prime picture pair (bird and puddle), the second participant (assigned to group 2) would see the no agent picture pair (book and puddle), and so on. Thus, a total of 28 experimental picture pairs and 112 filler picture pairs were shown to each participant.

Auditory stimuli. In addition to the 28 experimental verbs, 70 nonwords, 28 nouns, and 14 adjectives were used in the current experiments. Target words were matched for length and frequency, and were recorded by a male native English speaker at a sample rate of 44,100 Hz and a 16 bit sound resolution. Participants heard all words through headphones, and responded to them with a button box.

Procedure. Participants were seated at a comfortable distance from the computer monitor, with the button box placed directly in front of them. The participants were then told that they would see some images appear on the screen and, when they heard a word, they were to make a decision with the button box as to whether the word was real or not. A right button press was used to indicate 'yes' (the word was real), and a left button press was used to indicate a 'no' (the word was not real). Participants were also instructed as to the importance of responding as quickly and accurately as they could.

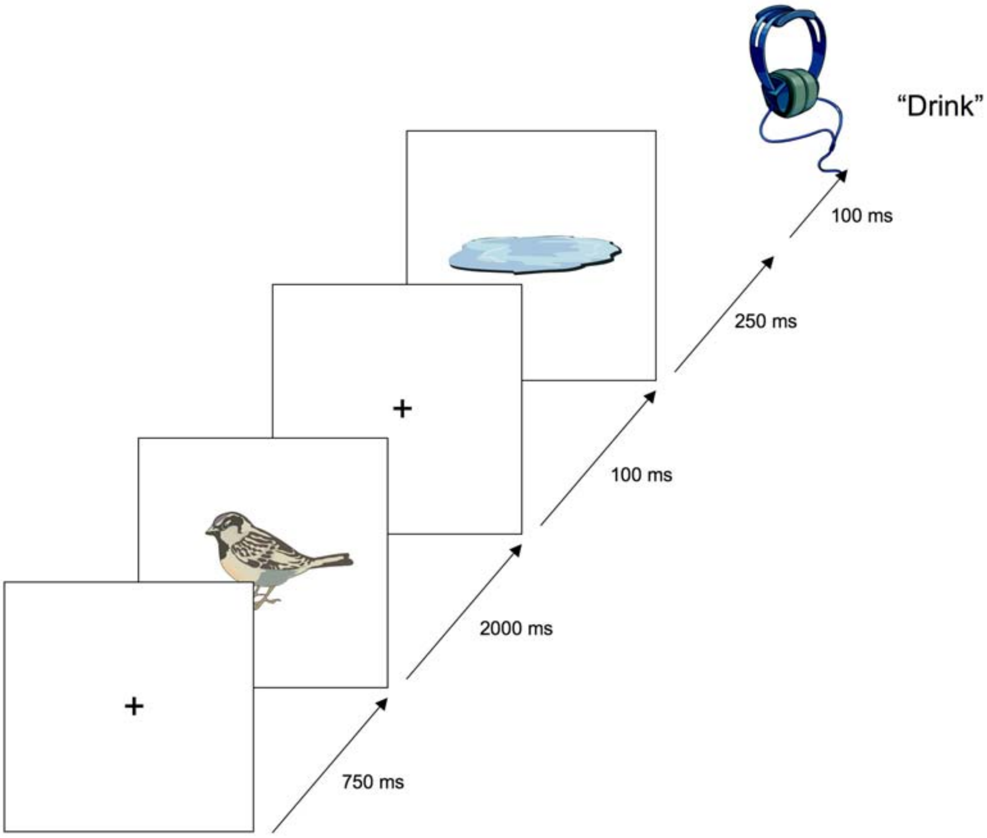


Fig. 2.3. *Schematic of the trial procedure for the target 'drink'.*

Figure 2.3 shows the procedure of an example trial, where +'s denote fixation crosses. Before each trial, the message 'Press Key' was displayed, and to move onto the next trial, either of the buttons on the box could be pressed. Any time this message was shown, participants could take a break if they needed one. On each trial, a fixation cross would appear first, and participants were required to fixate it. Then the two pictures would be shown, separated by another fixation cross. A short time after seeing the stimuli, participants would hear the word they were to make a decision on. There were 142 trials in total, including six practice trials, and the experiment lasted roughly 20 minutes.

Results

Table 2.1 shows the means and standard deviations for each of the four conditions, and figure 2.4 shows that, on average, participants responded fastest in the prime condition.

	Prime	No-agent	No-theme	Baseline
M	712	760	761	787
SD	146	191	219	197

Table 2.1. *Means and standard deviations for each condition, with second picture exposure durations of 250 ms.*

Analyses of response times for this and all subsequent experiments were conducted only on *correct* responses. A repeated measures ANOVA revealed a significant main effect of picture pair type (*F*1(3, 105) = 4.06, p = .009; *F*2[2] (3, 81) = 2.191, p = .113). Paired comparisons revealed significant differences between ‘prime’ and ‘baseline’ conditions (*F*1(35) = -2.85, p = .007; *F*2 (27) = -2.32, p = .028); ‘prime’ and ‘no agent’ conditions (*F*1(35) = -2.54, p = .016; *F*2(27) = -1.90, p = .068); and ‘prime’ and ‘no theme’ conditions (*F*1(35) = -2.44, p = .020; *F*2(27) = -2.26, p = .032).

---

[2] Where test statistics are numerically labelled, 1 refers to by-subjects analyses and 2 refers to by-items analyses.

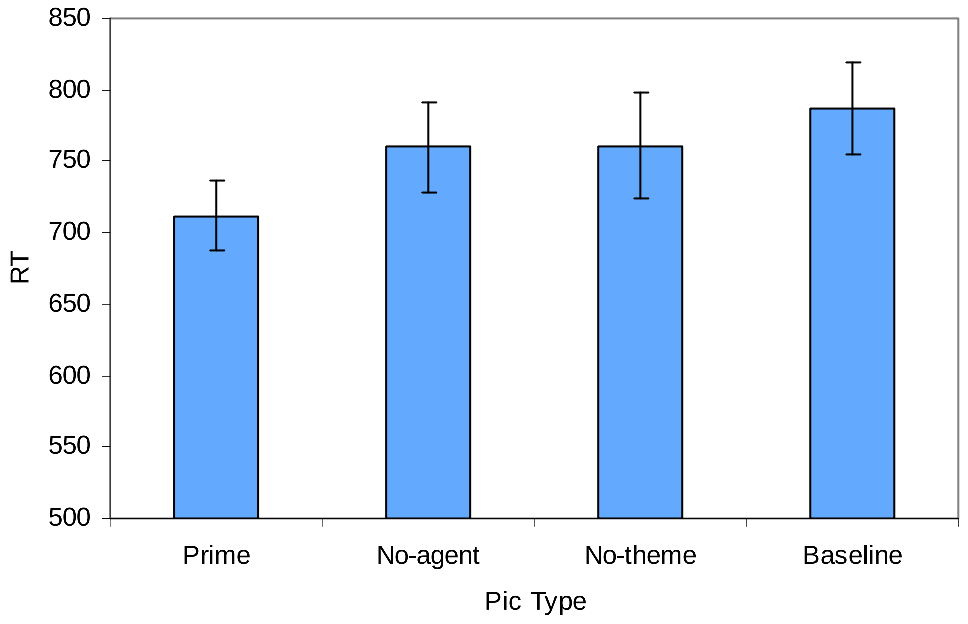


Fig. 2.4. *Mean response times at 250 ms, with standard error bars.*

Although participants responded faster in the ‘no agent’ than the ‘no theme’ condition, faster in the ‘no theme’ than the ‘baseline’ condition, and faster in the ‘no agent’ than the ‘baseline’ condition, these differences were not significant (all p>.05).

Participants’ accuracy in this and all subsequent experiments was measured in terms of the percentage of experimental trials in which the correct response was chosen (M = 89%, SD = 5%). To assess whether participants’ accuracy improved from the first half to the second half of trials, a one-way repeated measures ANOVA was conducted with a single factor of trial split (either the first or the second half). There was no difference in participants’ average number of correct responses between the first and second half of the experiment (p > .05).

In order to asses whether accuracy rates differed between condition, a repeated measures ANOVA was conducted with one factor of picture pair type (prime, no-agent, no-theme, or baseline), with a dependent variable of accuracy (number of correct responses across conditions). There was no main effect of picture pair type, and paired comparisons revealed no significant differences between conditions with respect to accuracy (all p > .05). Adjustments were made for all multiple comparisons using Bonferroni corrections.

Discussion

As predicted, participants were faster to respond in lexical decision when both presented pictures afforded the event denoted by the target word. To use a specific example for the event 'drink', faster responses were observed to the auditorially presented target verb "drink", when it was preceded by pictures of a bird and a puddle, the combination of which affords a drinking event. Less priming was observed when there was no suitable agent to carry out the event (e.g. a bird), no suitable theme with which the event could take place (e.g. a puddle), or unrelated baseline objects were presented (e.g. a book and a toothbrush). On average, response times in the prime condition were significantly faster compared to every other condition.

These results can be interpreted with respect to the abstraction and combining of affordances from visually presented objects and agents. On priming trials, the agent in each case affords carrying out the event, and the theme affords having the event take place. Importantly, this combining of semantic information occurs across both the visual and linguistic modalities.

Continuing with the same example, it is assumed that upon seeing a puddle that has been preceded by a bird, a representation of drinking is generated. This representation overlaps with the representation engendered by the target word ‘drink’, and lexical decision response times to the target are facilitated. This notion of overlapping representations is relevant to a single-code view of semantic access (Chase and Clark, 1972), and suggests that semantic information can be combined to form composite representations, regardless of the original source of the information (in this case words or pictures).

Although the current data provide a demonstration of combinatorial processes operating within a cross-modal semantic priming paradigm, it is unclear whether these effects can be attributed to automatic processes, or processes requiring strategic control. Faster, more implicit processing of visual stimuli is thought to be elicited at shorter SOAs of between 100-250 ms (Neely, 1991), while longer SOAs of 600-1000 ms are thought to elicit more controlled, strategic processing (Sachs et al., 2008).

With respect to strategic processing, however, an attempt was made to establish whether the number of correct responses participants were making, increased from the first to the second half of the experiment. An improvement in accuracy in the second half of the experiment may have implicated the use of strategies (such as mnemonics, see Corbett, 1977; Neisser and Kerr, 1973; and chapter 6 for more discussion) to complete the task. Although in the current experiment effects of improvement were not found, the same accuracy analyses will be conducted in the experiments that follow.

For now, The SOA in the following experiment is to be reduced in an attempt to establish the automaticity (or lack thereof) of the effects reported thus far.

Experiment 2

This experiment is essentially the same as Experiment 1, except that second picture exposure durations were shortened from 250 to 100 ms.

Method

Participants. 48 participants from the University of York student community took part in the experiment, and each person was given £2 or course credit for participating. As with experiment 1, all were native English speakers and had normal or corrected-to-normal vision.

Materials. The same materials used in experiment 1 were used for experiment 2.

Design. The same design as experiment 1 was used for experiment 2.

Procedure. The procedure was also the same as in experiment 1, except for second picture exposure durations of 100 rather than 250 ms.

Results

Table 2.2 shows the means and standard deviations for each of the four conditions with second picture exposure durations of 100 ms. Figure 2.5 shows the average response times for each condition.

	Prime	No-agent	No-theme	Baseline
M	717	717	746	756
SD	145	126	155	175

Table 2.2. *Means and standard deviations for each condition, with second picture exposure durations of 100 ms.*

Mauchly’s test revealed the assumption of sphericity had been violated ($\chi^2(5) = 23.83$, $p < .001$), so the following test was corrected using Greenhouse-Geisser estimates of sphericity. A repeated measures ANOVA revealed no significant main effect of picture type ($F1(2.34, 110.03) = 1.99$, $p = .134$; $F2(3, 81) = 2.594$, $p = .058$).

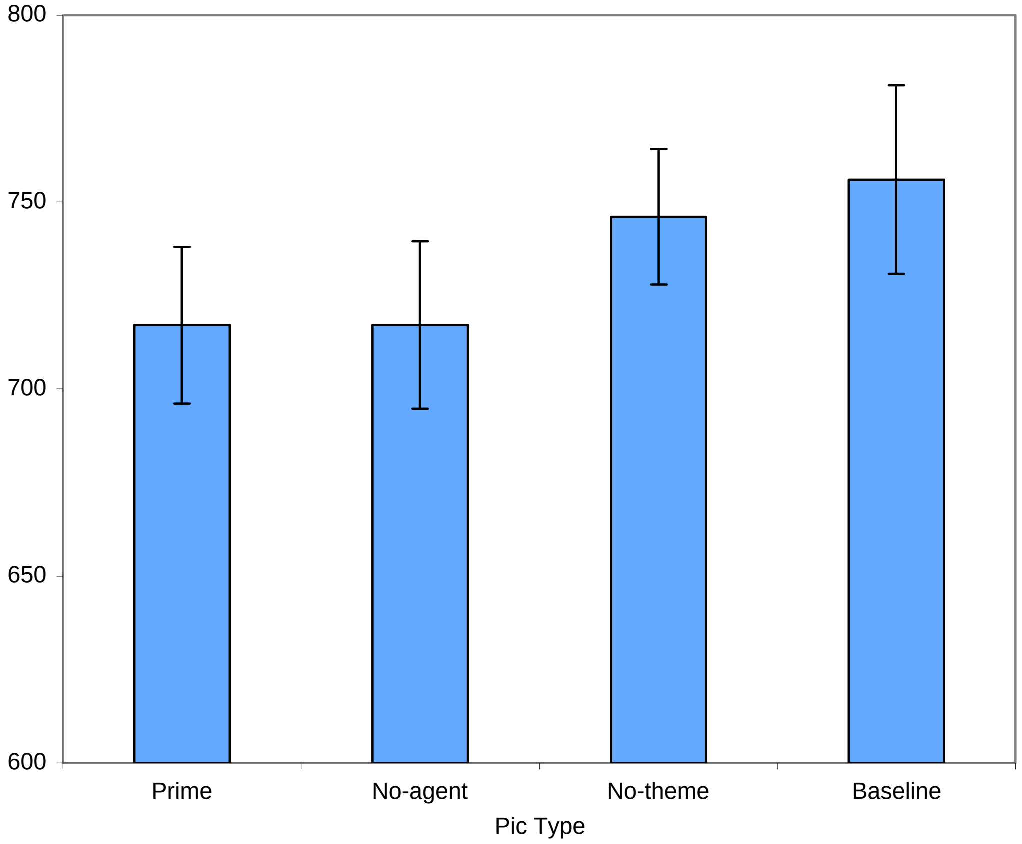


Fig. 2.5. *Mean response times at 100 ms, with standard error bars.*

Participants did not respond any slower or faster as a function of picture type. Paired comparisons revealed no significant differences in response times between the conditions of ‘prime’ and ‘baseline’, ‘prime’ and ‘no agent’, ‘prime’ and ‘no theme’, ‘no theme’ and ‘no agent’, ‘no theme’ and ‘baseline’, and finally, ‘no agent’ and ‘baseline’ (all p > .0.5).

Participants’ responses were, on average, very accurate (M = 90%, SD = 4%), and there was no difference in accuracy from the first to the second half of the experiment (p > .05). In order to establish the effect of these conditions on accuracy rates, a repeated measures ANOVA was conducted with one factor of picture pair type (prime, no-agent, no-theme, or baseline), with a dependent variable of accuracy (number of correct responses across conditions). There were no differences between conditions with respect to accuracy (all p > .05). Adjustments were made for all multiple comparisons using Bonferroni corrections.

General discussion

In experiment 1, participants' average response times to targets were significantly faster in the prime condition (where, for example, they might receive the bird and puddle items for the first and second pictures, respectively) compared to every other condition. In addition, there was a significant difference between average response times to prime and baseline items. When the exposure duration of second picture items was reduced from 250 to 100 ms in experiment 2, there were no significant differences in participants' average response times between any conditions.

The significant differences between the prime and all other conditions at 250 ms suggest that priming was not induced purely by one experimental item on its own, but by the combination of both. For example, for the target word ‘drink’, participants responded significantly faster when the bird was shown with the puddle (the ‘prime’ condition), than when either the bird was shown without the puddle (the ‘no-theme’ condition), or the puddle was shown without the bird (the ‘no-agent’ condition).

The results for experiment 2 suggest that at shorter exposure durations, there may not be enough time for participants to construct event representations from available item affordances. The nature of these experiments are such that any facilitation in responses to the target can only take effect once the affordances of items have been successfully combined. Therefore, as the event representation in this case can be thought of as a composite of combined item affordances, the additional time needed in these experiments for facilitation to the target word is a reflection of the additional information that needs to be combined.

There are three major implications of these results. Firstly, that these activated representations were rich enough to contain event-based information; secondly, that these combinatorial processes are likely to require more time to operate than those of gist abstraction; and thirdly, that semantic information is smoothly integrated from one picture to another, and is not simply replaced.

To take the first implication, these results tell us something of the nature of these activated representations in that they appear to code event-based information (the types of events a given item can engage in) along with commonly activated semantic features of the item(s) in question. Functional interactions between objects and their facilitation of task-dependent targets have been established before (e.g. Green & Hummel, 2006), but what these current experiments indicate is that representations can be activated based on the types of interaction afforded by the combination of two semantically unrelated items. For example, a bird and a puddle are not semantically related, but the affordances of each are combined only through the event that links them (drinking).

If one is to assume that recognition of these sequentially presented objects is necessary for these combinatorial processes to operate, then it is relevant that the processing requirements and time-course of gist abstraction appear to be different from those of object recognition. Loschky et al. (2007) argues that information abstracted in the form of gist is associated with holistic, low-level properties as apposed to information pertaining to the recognition of the individual objects comprising a scene. Abstracting meaning in terms of object type is more effortful than abstracting meaning in the form of gist (Dobel, 2007) and studies of gist abstraction emphasise the use of global spatial cues (such as layout of items) to establish the ‘identity’ of visual scenes (Henderson & Ferreira, 2004). Such cues are not available in the current set of experiments, and an ‘identity’ may be more difficult to attach to two objects linked only by the nature of their possible interactions with one another. In the current set of experiments, in order for the affordances of objects to be combined, the objects have to be recognised first. It is unlikely that without this crucial step these effects are unable to operate on our lexical decision on the target word.

The last implication is that semantic information from one picture does not simply replace that of the previous one but they are combined via event-based knowledge. Indeed, our experience of the visual world is not one of discrete, self-contained packets of information, but of fluid, seamless comprehension of our surroundings (Irwin, 1991). However, studies of information integration from one saccadic eye movement to the next suggest that although we rapidly combine separate sources of visual information, only a few representations are utilised at any one time, before being replaced by a different set of representations (Irwin, 1991). Of course, if semantic information pertaining to a previously seen item is replaced when a second saccade is executed, it is unclear how the combining of semantic information between objects in our visual environment would occur in the first place.

Experiments serving to both establish and support semantic combinatorial effects in the linguistic domain are numerous (e.g. Tabossi, 1988; Glenberg &

Robertson, 2000; Andrews et al., 1993; Seidenberg et al., 1982; Potter & Faulconer, 1979). However, the current cross-modal paradigm may be an inconclusive demonstration of a similar effect operating in the visual domain. These experiments demonstrate priming based on overlapping representations constructed from the semantics inherent in the presented pictures and the semantics engendered by the target verb. In order to establish whether the combinatorial effects demonstrated in linguistic paradigms operate to the same degree in the visual domain, reliance on the LDT must be altered in favour of a purely visual task. Given that combinatorial effects have been established in the linguistic domain, the operation of these effects in a paradigm that utilises only visual information would indicate that these effects are not induced by linguistic stimuli alone, but are elicited to the same extent by visual objects. This is to be a focus of the experiments presented in the next chapter. These subsequent experiments shall also provide an investigation of the causal nature of these relationships between presented items.

Chapter 3

The present chapter describes three experiments that build upon the findings presented in the previous chapter. It begins with an overview of those findings and their theoretical implications, and follows with the motivation for the new experiments. A novel paradigm will be presented and discussed that explores combinatory semantic processing through the priming of visual object recognition using a non-linguistic task.

In experiment 1, participants were faster to respond to the word ‘drink’ in a lexical decision task if it had been preceded by a picture of a bird (the agent) and then a picture of a puddle (the theme). Crucially, if either the agent was shown without the theme or the theme was shown without the agent, response times were much slower. The verb ‘drink’ in this case is presumed to be primed by the combination of affordances derived from the bird and the puddle. The bird affords drinking from the puddle and the puddle affords being drunk from. These combined affordances are used to generate an event representation that corresponds to these abstracted affordances. This event representation overlaps with the conceptual structure that would ordinarily be activated on hearing the target ‘drink’. Priming in this case arises from some information pertaining to drinking becoming pre-activated from the prior generated event representation. This priming from the constructed event representation then contributes to the observed facilitation of response times to the verb.

These findings offer support for the notion of a visual combinatorial process, and expand upon findings of a similar nature in the linguistic domain (e.g. Tabossi (1988), Potter & Faulconer (1979), see the ‘priming effects’ subheading in the introduction for discussion). These studies reveal the importance of context in language processing and provide examples of the modulation of response times as a function of the amount of information that can be combined in sentential events. A similar process operating on both linguistic and visual stimuli implicates a mechanism common to both domains. The combining of information to form composite representations thus becomes a cognitive function fundamental to our comprehension of our surroundings, regardless of the source from which it comes.

There are, however, two unresolved issues from experiments 1 and 2 that provide part of the motivation for the current experiment. First, these findings tell us nothing of the *nature* of the representations activated: Are the event-representations coded in such a way as to include the specific agents and themes seen by the participant (thereby encoding the specific events afforded by those specific items), or are they more abstract, coded in terms of a general instance of the same event (that may be more independent of the specific items)? Second, the use of the lexical decision task is problematic for an account of a combinatorial effect arising purely from visual perception alone. Establishing the occurrence of this combinatorial process in the visual domain is essential if it is to be maintained that this effect is an integral part of our processing of the external environment.

In addition to addressing the aforementioned issues concerning abstract representation and the use of language, the current paradigm allows deeper investigation of the *causal* structure of an event, and what the representations likely contain in terms of causal information. Causal structure can be thought of in this case as the identifiable segments of a causal chain and how these segments are linked; more simply, a ‘who did what to whom’ sequence. Each of these points shall be discussed in turn, leading to a rationale for the current experiment that aims to address these issues.

To take the first point, with respect to the nature of the activated representations there are two accounts of the data for experiment 1. Continuing with the ‘bird-puddle’ example introduced in the previous chapter, participants’ response times to the word ‘drink’ are either speeded because the word matches an activated representation of the bird drinking from the puddle, or because the activated representation concerns a general act of drinking (which could include, for example, drinking from a glass), and it is this that matches the semantics engendered by the word.

Colcombe and Wyer (2002) argue that generalised representations of an event may be activated if that event has been experienced repeatedly. In other words, for any given event, if one has extensive experiential knowledge of it, then the representations activated upon being confronted with it again, are likely to be more abstract. Similarly, Zacks and Tversky (2001) assert that a schema or ‘script’ for a particular event may be created by repeated exposure to that particular type of event. These abstract representations do not contain information concerning what or who the participants in a particular event are, where the event takes place, or the amount of time the event uses from its inception to its completion.

A demonstration of the nature of these representations with respect to their level of generality could proceed as follows: If, in experiment 1, participants were constructing abstract representations of the intended events, then an appropriate test for this would be to have participants make judgements on targets that are associated with an interpretation of the event that is *not* afforded by the objects in the scene. For example, if upon seeing a bird (the agent) and then a puddle (the theme), an abstract drinking representation is activated, response times to a concept (such as a glass) which affords a drinking event but not the specific kind of drinking that a bird would engage in, would be facilitated compared to a control. An appropriate control condition for these items would be something that is unlikely to interact with the theme (such as a boy for the bird and puddle example). The study described below employs just such items.

With respect to the use of lexical decision to probe the activation of these event-representations, the notion of a combinatorial semantic process introduced in the first chapter was presented on the basis that it represents an operation common not only to the linguistic domain, but to the visual domain also. However, the paradigm described in the previous chapter still utilised language as a basis for collecting response times and this becomes problematic for an account that aims to provide evidence for such an effect operating during visual processing and object recognition.

The literature is replete with examples of two-way interplay between vision and language, particularly within the field of psycholinguistics (Henderson & Ferreira, 2004; Glaser, 1992). The mapping of language onto a visual event has been demonstrated with studies utilising the visual world paradigm (Altmann & Kamide, 1999; Kamide, Altmann & Haywood, 2003; Tanenhaus et al., 1995), and words and pictures have been shown to access a single semantic store for both types of information (Vanderwart, 1984; Chase & Clark, 1972). Given that language is seen to have such a facilitatory impact on object recognition and visual perception, its elimination from the current paradigm is essential in order to establish the occurrence of uncontaminated combinatorial processes operating on visual stimuli alone.

The current experiment will address this issue by using an artifact judgement task instead of a lexical decision task. The judgement is essentially a semantic categorisation task in which the participant is asked to make a decision on a target as to whether it is a man-made object or not (see Del'Acqua & Grainger, 1999). Participants will make this judgment on a visual target which represents an abstract instantiation of a particular event. So, for the bird and puddle example, participants would first be presented with a picture of a bird and a puddle and then they would be presented with a picture of a glass, upon which they will be required to make their artifact decision (see figure 3.1). The control for this example would consist of, for the first picture, a *boy* and a puddle, and for the second picture, the same target consisting of a glass.

To continue with the same example, it is assumed that priming of the glass from a bird and a puddle will be due to the generation of a general representation of a drinking event from the combining of affordances from the first picture. The bird and the puddle in this case affords drinking, but not from a glass specifically. The glass would only be primed if the representation used to recognise the target encodes the event as an abstract instantiation of drinking in general, that encompasses any form of drinking, including from puddles and glasses.

With respect to the methodology for this experiment, the artifact judgement task is used as an alternative to the lexical decision due to their relative similarity - the only main difference between the two is that the former requires participants to make a decision on visual instead of linguistic stimuli. In addition, this task has been shown to be sensitive to priming by pictorial stimuli, and does not require explicit attention to the first presented picture.

Prime condition

Picture 1

Picture 2

Control condition

Picture 1

Picture 2

Fig. 3.1. *An example of the stimuli for the event 'drink'.*

A final point to be discussed in advance of the study to be described below concerns the causal properties of visual events and the ways in which this information may be coded in the representations discussed in this section. In addition to the items used for experiment 1 (Henceforth referred to as 'bird-puddle' items), new items will be used within the same paradigm, which are endowed with causal information. The selection of these items is essentially capitalising on the fact that events usually involve a change in state in one or more of the constituent objects in a visual scene.

For these items (henceforth referred to as ‘knife-tyre’ items), participants will be presented with, firstly, a picture consisting of an object and an instrument that is capable of inducing a change in that object, for example, a knife and a tyre. Participants shall then be presented with the same object in a changed state, such as a burst tyre, and are required to make their artifact judgements on this target object (see appendix 2 for a list of bird-puddle and knife-tyre items). The control for these items shall consist of the same target, preceded by a picture of the target in its unchanged state and an instrument which is not capable of changing the target in any way (such as a ruler) (see figure 3.2).

Whereas the bird-puddle items do not have targets which are causally related to the prime (the glass, for example, is not associated in any way with the bird and puddle), the knife-tyre items have a causal structure: The tyre in the target picture target picture is burst *because* the previously presented knife and intact tyre allows bursting to take place. Therefore, within a single paradigm, the current experiment tests for a more generic event representation in the ‘glass target’ case, and a more object-specific event representation in the ‘burst tyre’ case.

Prime condition

Picture 1 Picture 2

Control condition

Picture 1 Picture 2

Fig. 3.2. *An example of stimuli for the event 'burst'.*

According to Zacks & Tversky (2001), a key feature of the processing of an event's structure is the perception of causality inherent in the event itself. These authors use a real-world account of the perception of event structure to discuss the fluid interplay of bottom-up and top-down processes operating during the processing of events and the activation of representations of those events. The use of these 'causal' items, below, is predicated on them providing a demonstration of the interaction during object recognition of bottom-up processes and top-down cognitive processing of perceptually significant configurations of objects.

If one is to expect priming from cases where objects have been paired with instruments capable of causing a change in that object, one must acknowledge the relative contributions of both bottom-up and top-down processing in the comprehension of a given visual event. The items that will be used to explore causality within visual events are distinct from the items used for experiments 1 and 2, as these items provide cues as to the event's causal structure.

In a sense, it was an incomplete event that was presented to participants responding in experiments 1 and 2: there was no perceptible change in the target that could have been precipitated by the agent. In the current paradigm however, if participants are presented with a knife and a tyre, the causal structure of the event is constrained so as to allow very few resultant states in the target – the most plausible of which is the bursting of the tyre. Here, participants are able to make use of processing of the causal structure in the events presented to them: the last 'segment' of the causal structure of the event (for example, the burst tyre) is included in this case. This utilisation of causal event structure influences object recognition so that the processing is consonant with the constraints of the affordances of the objects in the first presented picture.

Thus far, combinatorial effects have been demonstrated only in cases where an agent is depicted, a likely candidate for carrying out the event in question on each priming trial. However a crucial difference between these experiments and the knife-tyre portion of the current experiment is that for the latter, no agents will be depicted with the instruments and objects in first picture presentations. For events to be presented with inherent casual structure information (e.g. the knife bursts the tyre and the target is depicted as an object in its end-state as a result of this bursting), a likely agent is not a necessary inclusion. For example, we know that, given a knife and a tyre, the tyre affords being burst, regardless of whether an agent is included in the scene. Furthermore, studies have demonstrated sensitivity to functional groupings of objects without the need of agents in visual scenes (e.g. Green and Hummel, 2006). It may be that in these studies, an agent's involvement is implicit in the casual nature of event. The likelihood of there being a causal agent in a given event may be so high that its explicit inclusion is not necessary.

So to reiterate the basic design of the current paradigm, there are two sets of items, one set seeking to address the distinction between abstract and concrete representations, and one seeking to investigate participants' utilisation of causal structure information in their processing of visual events. For this second set of items, causal information is derived from instruments and objects (such as a knife and tyre), based on the event these two items, together, afford (in this case, bursting). This priming of causal structure is then expected to facilitate responses to a target which corresponds to the resultant state of this causal event (such as a burst tyre). Both sets of items shall address the language task issue as for both sets, participants make an artifact judgement on the target, rather than a lexical decision. The current paradigm, although discussed briefly here, will be described in detail in the following 'methods' section.

The bird-puddle items were included to tease apart the distinction between more general abstract representations, and the more concrete representations involving the constituent objects directly. However, the exploratory nature of this early experiment makes generating specific predictions concerning the bird-puddle items difficult. Either null or significant results in this case would reveal insights into the nature of these representations: If there is no significant difference between conditions then it is likely that the representations activated in these paradigms contain information concerning only the affordances and potential interactions of the agents and themes presented in the first picture. If the opposite is the case, then support can be provided for a more abstract representation activated upon seeing certain combinations of objects. However, although no firm conclusions can generally be drawn from a null result, we shall, in the next chapter, address this issue and explore further this notion of object-specific representations using a modified paradigm.

Generating predictions for the knife-tyre items is much less exacting because the combinatorial arguments raised in this and the previous chapters are only supported with one outcome. It is predicted, therefore, that participants will be faster to respond in artifact judgement to a target in a changed state (e.g. a burst tyre), when it has been preceded by that item in its unchanged state and an instrument which is capable of bringing about this change in the target (e.g. an intact tyre and a knife). In terms of the processes that are assumed operate here, the affordances of the object and instrument are initially combined to form an event representation based on the event that, together, these items allow. This representation then primes the likely change in state based on the causal structure of the event. It is this priming of the target that would result in the facilitation of response times to this object. Although this can be taken as the principal account of the combinatorial effects described thus far, there is an alternative account based on integrative processes at the target, which is to be discussed in detail later.

The present section has provided a rationale for the current experiment based on findings from the previous experiment and also a desire to expand the scope of this work. Firstly, the current experiment shall investigate the nature of event representation with bird-puddle targets consisting of an object related to the abstract conception of the presented event (e.g. a glass for the event ‘drink’). The use of language shall be eradicated from the current experiment with the use of an artifact judgement task, and the causal nature of events shall also be exploited with the use of targets in changed states for the knife-tyre items.

Experiment 3

Method

Participants. 40 University of York students participated in the experiment for half-an-hour’s course credit or £2. All participants had normal or corrected-to-normal vision.

Materials.

Bird-puddle items: Experimental items. Figure 3.1 shows an example of the stimuli used for the ‘drinking’ event within the bird-puddle items. The first picture in the prime condition consisted of an agent (e.g. a bird) with its theme (e.g. a puddle) and the second picture in the prime condition consisted of the target (e.g. a glass). The first picture in the control condition consisted of a different agent (e.g. a boy) with the same theme (e.g. a puddle) and the second picture in the control condition consisted of the same target (e.g. a glass).

There were 18 experimental bird-puddle items in total, and all targets for experimental items consisted of man-made objects. Participants saw all experimental items but primes and controls were rotated across lists. For example, participants assigned to list 1 saw 9 primes and 9 controls. Participants assigned to list 2 saw the other 9 primes and 9 controls. In other words, a prime in list 1 became a control in list 2 and so on.

Bird-puddle items: Fillers. Filler items for the bird-puddle set were designed so as to correspond to the experimental items. For example, 9 fillers were related objects, with animate entities (mirroring the 9 priming items seen in a given list) and the other 9 were unrelated objects with animate entities (mirroring the 9 control items seen in a given list). Related objects might include things such as a policeman and handcuffs or an infant and a dummy, while unrelated objects might include an actress and a stapler or a pelican and a rugby ball. There were 18 bird-puddle fillers in total and each 'target' consisted of natural objects. The 18 knife-tyre experimental items and the 18 knife-tyre fillers were treated as fillers for the bird-puddle items. So in effect, there were 54 bird-puddle fillers in total.

Knife-tyre items: Experimental items. Figure 3.2 shows an example of the type of stimuli used for the knife-tyre experimental items. For the prime condition, the first picture consisted of an object in its unchanged state (e.g. an intact tyre) and an instrument (e.g. a knife), and the second picture consisted of the prior object in its changed state (e.g. a burst tyre). For the control condition, the first picture consisted of an object in its unchanged state (e.g. an intact tyre) and a different instrument (e.g. a ruler), and the second picture consisted of the same target (e.g. a burst tyre). These control objects were also designed to be similar in shape and size to the priming objects. There were 18 experimental knife-tyre items in total and each target was a man-made object.

Knife-tyre items: Fillers. For the knife-tyre set, filler items also corresponded to experimental items in that they consisted of 9 related objects (mirroring the 9 prime items seen by participants in a given list) and 9 unrelated objects (mirroring the 9 control items seen by participants in a given list). Related objects in this case might include things such as a table and a chair or a bowl and spoon, while unrelated objects might include a sofa and a rake or a computer and a spade. So there were 18 knife-tyre fillers in total and each ‘target’ consisted of a natural object. Just as with the other set of items, the 18 bird-puddle fillers and the 18 bird-puddle experimental items formed a total of 54 fillers for these knife-tyre items.

Participants were assigned either to list 1 or list 2. For the experimental items in list one, half of them were primes, and the other half were controls. For the experimental items in list 2, each prime for list 1 was a control in list 2, and vice versa. So for example, participant 1 may have seen a prime for the first trial and a control for the second, while participant 2 may have seen a control for the first trial and a prime for the second. All items were downloaded from www.clipart.com and were shown at a resolution of 640×480 on the 17” monitor of an Apple iMac computer.

Norming. The items for bird-puddle and knife-tyre sets were normed using online questionnaires (see appendix 6 for instructions). These were designed to test for things such as the plausibility of events by asking participants how likely a particular event is to occur, given the combination of items provided. Participants made their choices on a scale of 1 to 7, with 7 being 'highly likely' and 1 being 'highly unlikely'. 21 different participants participated in answering the questions from each 'bird-puddle' questionnaire. The data from the norming studies will be described in the Results section below.

There were three different questionnaires for the bird-puddle items, each dealing with a different question, and participants were randomly assigned to one of these different versions. For Version 1, participants were required to make a choice as to how likely they thought an event (e.g. drinking) was to occur given a combination of objects (e.g. a bird and a puddle). This question was asked of participants so as to gauge how effective experimental items were in affording a particular event. If participants regarded a particular event to be 'highly unlikely' given a priming combination of objects, then it could be assumed that these items would not induce the activation of corresponding representations when presented to them in the semantic priming paradigm.

For Version 2, participants were asked to indicate how likely they regarded an event to happen (e.g. drinking) given one of the targets (e.g. a glass). This question was necessary for the selection of appropriate targets. Again, if participants regarded events to be highly unlikely given a particular target, then it could be that these items would not effectively afford the event in question and so would need to be eliminated from the item set.

For Version 3, participants were asked to rate how likely an event would be to occur (e.g. drinking) given an agent (e.g. a bird). This question was asked in order to prevent the inclusion of items with which the representation of events could be activated by the agent alone. If priming were driven by the agent alone, then this would confound an account of such effects in terms of the *combination* of items in the first picture.

Norming questionnaires for the knife-tyre items were also split by several questions. Version 1 sought to establish how plausible a given event was, given a particular combination of items. For example, they might have been presented with a knife and a tyre and they would be required to make a choice on a scale of 1 to 7 as to how likely they thought that a particular event would occur in everyday life. As with the bird-puddle questionnaires, a choice of 7 indicated an event was 'highly likely' to occur, while a choice of 1 indicated an event was 'highly unlikely' to occur.

Version 2 for the knife-tyre items asked participants to rate how likely it is that a particular instrument (e.g. a knife) caused the state of a particular object (e.g. a burst tyre). This needs to be established because if participants are not sensitive to the affordances of the instrument then it is unlikely a representation involving the bursting of the tyre will be activated. As trials in the control condition for knife-tyre items consisted of different instruments (e.g. a ruler instead of a knife), half of the items on version 2 were priming combinations and the other half were control combinations. Items that were priming combinations on version 2 became control conditions on version 3, and vice versa.

Version 3 was designed to establish the commonality of knife-tyre targets. Participants receiving this questionnaire were asked to rate how common they thought a target object was in everyday life. Again, these ratings were to be selected from a scale of 1 to 7, with 7 denoting a ‘highly common’ item and a score of 1 denoting a ‘highly uncommon’ item. It is assumed that the commonality of targets will have an impact on their recognition, regardless of any event representations activated. 21 different participants participated in answering the questions from each questionnaire.

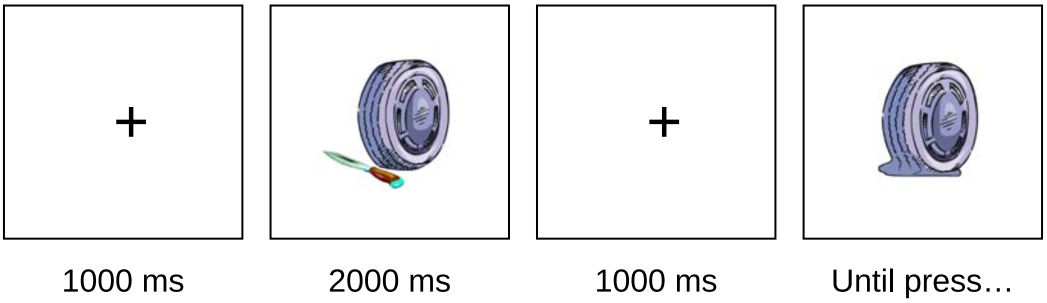
Procedure. Participants were seated at a comfortable distance from the computer monitor, with a button box placed directly in front of them. The participants were asked to sign consent forms and were then given onscreen instructions (see appendix 3)[3]. The instructions informed participants that they would see two images in each trial, and that when they see the second image they were to make a judgment as to whether they think it consists of something which is man-made or not man-made. They were to make this decision with the button box, and they were instructed to respond as quickly and accurately as possible. A right button press was used to indicate ‘yes’ (the second image is man-made), and a left button press was used to indicate a ‘no’ (the second image is not man-made).

Figure 3.3 shows a schematic of the trial procedure. First, a fixation cross appeared on the screen for 1000 ms. Then, the first picture appeared for 2000 ms followed by another fixation cross for 1000 ms. Finally, the second picture appeared and remained on the screen until one of the buttons on the button box was pressed.

---

[3] The instructions required participants to concentrate on all items as there would be a short memory test when the experiment had finished. Upon debriefing, however, participants were informed that there would be no memory test, but that they were told this because the experimenter was wary of them simply ignoring the first picture.

Prime condition



Control condition

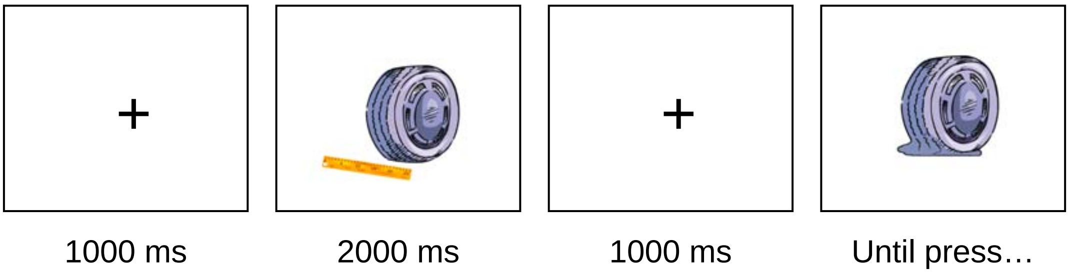


Fig. 3.3. *An example of the procedure for each trial.*

Before each trial, the message 'Press Key' was displayed, and to move onto the next trial, either of the buttons on the box could be pressed. Any time this message was shown, participants could take a break if they needed one. When the experiment had ended, participants were presented with debriefing information (see appendix 4.) Before participants began the experiment, they were presented with 4 practice trials. There were 72 trials in total, and the experiment lasted roughly 30 minutes.

Results

Bird-puddle Priming. Figure 3.4 shows that, on average, participants responded faster in the control condition (M = 1087, SD = 426) than the prime condition (M = 1143, SD = 461), and this difference was significant (*F*1(39) = 2.174, p = .036; *F*2(17) = 1.092, p = .290). Response times more than three standard deviations above the mean were replaced with the mean plus three standard deviations, and response times more than three standard deviations below the mean were replaced with the mean minus three standard deviations. However, altering this trimming from 3 to 4 standard deviations eradicates the difference between prime and control conditions (*F*1(39) = -.320, p = .077; *F*2(17) = -1.642, p = .119).

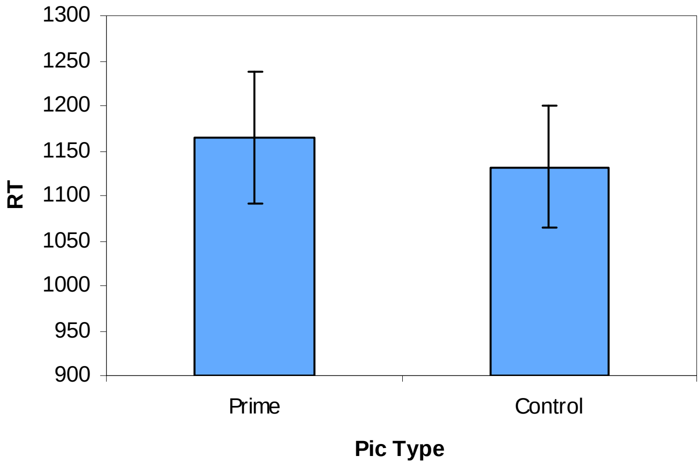


Fig. 3.4. *Mean 'bird-puddle' response times for both conditions, with standard error bars.*

Participants' responses were very accurate on average (M = 95%, SD = 5%) and there was no difference of average number of correct responses between the first and second halves of the experiment (p > .05). There was no difference between conditions with respect to accuracy (p > .05).

Bird-puddle Norming. Table 3.1 shows the means and standard deviations for ratings for each of the three different types of norming. The mean figure for ‘Plaus (prime)’ represents the average rating out of 7 given to prime items for how likely an event (such as drinking) is to occur. The mean figure for ‘Target’ represents the average rating score out of 7 given to prime targets for how likely an event is to happen. Finally, the mean figure for ‘Agent’ represents the average rating out of 7 for prime agents for how likely an event is to happen.

	Plaus (prime)	Target	Agent
Mean	5.64	5.62	4.02
SD	0.83	0.93	1.34

Table 3.1. *Means and standard deviations for each of the three norming types.*

Table 3.2 shows Pearson’s correlations between response times in the priming condition and each of the rating types. Negative correlations between response times and both plausibility ratings and agent ratings indicate that the more likely an event or the more likely an agent is to cause an event, the faster the response times. The only positive correlation in the table is that between response times for the prime condition and the ratings for targets. This indicates that the more likely an event is to happen given a target, the slower the response times are to that item. None of these correlations, however, are significant (all $p > .05$).

	Plaus (prime)	Target	Agent
Prime RTs	-0.175	0.246	-0.084
Plaus (prime)	-------	-0.359	-0.015
Target	-------	-------	-0.214
Agent	-------	-------	-------

Table 3.2. *Pearson's r values for correlations between bird-puddle prime condition response times and each rating type.*

Knife-tyre priming. Figure 3.5 shows that, on average, participants responded faster in the prime condition (M = 1102, SD = 435) than in the control condition (M = 1189, SD = 488): (*F*1(39) = -2.843, p = .007; *F*2(17) = -1.640, p = .119).

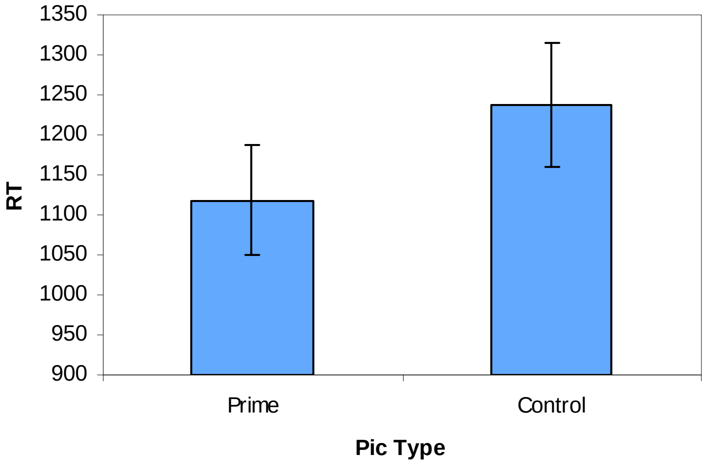


Fig. 3.5. *Mean 'knife-tyre' response times for both conditions, with standard error bars.*

Again, participants' responses were very accurate on average (M = 90%, SD = 4%) and there was no difference of average number of correct responses between the first and second halves of the experiment (p > .05). There was no difference between conditions with respect to accuracy (p > .05).

Knife-tyre Norming. Table 3.3 shows the means and standard deviations for ratings for each of the three different types of knife-tyre norming. The mean figure for ‘Plaus (prime)’ represents the average rating out of 7 given to prime items for how plausible a given event is regarded to be. The mean figure for ‘Plaus (control)’ represents the average rating score out of 7 given to control items for how plausible an event is. Lastly, the mean figure for ‘Commonality’ represents the average rating out of 7 given to targets in response to how common these items are deemed to be. The mean difference between plausibility ratings for prime and control items is significant ($F(17) = 14.627$, $p<.001$), indicating that given events would be more likely to occur if a priming combination was presented, compared to when a control combination was presented.

	Plaus (prime)	Plaus (control)	Commonality
Mean	4.92	1.84	3.97
SD	0.83	0.69	1.13

Table 3.3. *Means and standard deviations for each of the three norming types.*

Table 3.4 shows Pearson’s correlations between response times in the priming and control conditions and each of the rating types. The only significant r value was between commonality and plausibility (prime) ratings. This positive correlation indicates that the more plausible an event in the prime condition, the more common its corresponding target is likely to be.

	Plaus (prime)	Plaus (control)	Commonality
Prime RTs	-0.09	0.08	-0.07
Control RTs	-0.24	-0.07	-0.05
Plaus (prime)	-------	-0.17	0.61*
Plaus (control)	-------	-------	-0.08
Commonality	-------	-------	-------

*p<.05

Table 3.4. *Pearson's r values for correlations between knife-tyre prime and control condition response times and each rating type.*

Correlations were also computed to establish whether the magnitude of the difference in RTs between the prime and control conditions was predicted by the magnitude in difference between rating scores for these conditions. A Pearson's correlation between these differences was not significant, however (r = 0.04, p = 0.866).

Discussion

For the bird-puddle items in this experiment, participants were significantly faster to respond in the control condition (e.g. picture 1: boy and puddle, picture 2: glass), than they were to respond in the prime condition (e.g. picture 1: bird and puddle, picture 2: glass). The fact that results from this half of the stimuli ran contrary to the prediction may have been due to a flaw in the study concerning the items chosen to delineate this distinction between concrete and abstract event representations. Some of the control items were not appropriate examples of agents that do not afford interaction with presented objects. For example, although a boy is less likely to drink from a glass than an adult, it is more likely to drink from one than a bird. Perhaps a more appropriate design would be to include another condition to which this control could be compared: for example, a man for the first picture (an agent who is the most likely to drink from the glass compared to the others) and a glass for the second.

An indication that these results may be inconclusive comes from the fact that this significant effect of controls over primes disappears with a more conservative removal of outliers. Although this result appears to be unconvincing, a following experiment with more appropriate controls will be presented in a following chapter. This subsequent experiment will return to this issue of concrete and abstract representations with a stimulus set similar to the one used for the second half of the current stimulus set: the knife-tyre items.

For the knife-tyre items, participants were significantly faster to respond in the prime condition (picture 1: knife and tyre, picture 2: burst tyre), compared to the control condition (picture 1: knife and ruler, picture 2: burst tyre). Priming to the target arises here from an activated representation of the event, depicted in its end-state by the target. This representation is constructed from the combined affordances of the instrument and object. More specifically, if participants are presented with a knife and a tyre the affordances of these items are combined to form an event representation based on the event that, together, these items afford: knives afford bursting and tyres afford being burst. This representation then primes the likely change in state based on the causal structure of the event. This priming of the target results in the observed facilitation of response times to this object.

An important theoretical implication of the results for knife-tyre items concerns the aforementioned language-use issue raised from findings reported in the previous chapter. An advantage of the paradigm used for this experiment was that it was free from any potential influence of language and the data reported in this chapter suggest the event representations were activated purely from the visual information inherent in the stimuli presented to participants. This is significant because although several studies suggest that semantic information is readily combined in the linguistic domain (see Tabossi, 1988, Potter & Faulconer, 1979, Andrews et al.,1993), the current set of experiments aim to demonstrate a similar process operating in the visual domain.

It is important to mention, however, that although the paradigm contained no instances of language use, the participants were not prevented from labelling the stimuli themselves. In addition, the knife-tyre results from the current paradigm do not necessarily indicate that combinatorial effects are automatic processes of scene recognition and comprehension. An effective way to constrain the paradigm and induce automatic process is to reduce the exposure duration of the presented stimuli (Neely, 1991). The last experiment in this chapter will investigate this directly, in a paradigm where first picture exposure durations are shortened to 250 ms.

A potential criticism of the items selected for the knife-tyre part of the experiment is that in some of the experimental trials, there may have been a semantic association between the instrument and the target. To take the example of the items for the ‘open’ event, participants were shown a picture of an open can, preceded by a picture of a closed can with a can opener. These items are problematic because the can opener is ‘for’ opening cans and it may be the case that participants are not combining the affordances of the two objects here, but are primed to the appearance of an open can because of the clear function of the can opener. As Adamo & Ferber (2009) note, representations consonant with an instrument's affordance are activated upon seeing that instrument. These authors go further by proposing that, upon seeing an instrument, a representation is generated in visual short-term memory, encompassing its affordances and the types of 'surfaces' it can affect.

In order to eliminate this confound, the subsequent experiment shall use the same paradigm, but with these potentially problematic items removed from the stimulus set. In addition to these items being replaced by more suitable stimuli, two extra items were created, making the total number of experimental items 20. If these effects are being driven by such relationships between instruments and targets, then they should not be replicated in an experiment with more carefully controlled stimuli.

Experiment 4

Any items where a relationship between instruments and targets could be established were removed in this experiment, and replaced with items free of this potential confound. 11 items were replaced based on judgements of instrument-object association. There was a clear distinction between suitable and problematic items, such that the stimulus set could readily be split based on assumptions of association. These new items were also normed for plausibility of events; the likelihood of instruments causing changes in targets; and the commonality of target items.

Method

Participants. 30 University of York students participated in the experiment for half-an-hour's course credit or £2. All participants had normal or corrected-to-normal vision.

Materials. 11 of the knife-tyre items used in experiment 3 were replaced (see appendix 5). Two additional items were included with these replaced items, making 20 items in total for this experiment. The fillers were identical to experiment 3, except for two additional related filler items (corresponding to the two new experimental items in the set).

As before, all participants saw each item, but items were rotated across two lists. For example, a participant presented with items from list 1 would see 10 primes and 10 controls, and another participant presented with list 2 items would see the other 10 primes and the other 10 controls.

Norming. Every item presented to participants in experiment 4 was normed using online questionnaires (see appendix 7 for instructions). There were three versions of these questionnaires, with each one asking a specific question concerning the items used for experiment 2 These questionnaires differed from those used in the previous experiment, as these questions focussed on events in terms of instruments and objects, as apposed to agents and objects.

Version 1 concerned the plausibility of the events depicted by the items. For this questionnaire, participants were shown two pictures for each question, representing a certain event. They were then asked to make a decision on a scale of one-to-seven as to how likely they thought the depicted event was to occur in everyday life, with ‘1’ denoting highly unlikely and ‘7’ denoting highly likely. For example, participants may have been shown a tyre and a knife for the first picture, and a burst tyre for the second. They were then asked to decide on the scale how likely they thought the bursting of the tyre would be to happen in everyday life. These questions were not presented to the participants in text (each ‘question’ appeared as two pictures above the 7-point scale), but they were required simply to make the same decision for pair of pictures.

Version 2 of the questionnaire concerned the likelihood of the instrument in each case causing the change in the target. Here the same scale was used, and again, a choice of ‘1’ represented highly unlikely, and ‘7’ highly likely. For example, participants may have been shown a burst tyre for the first picture, and a knife for the second. In this case, their decision should have been based on how likely they thought it was that the tyre was burst by the knife.

Version 3 concerned the commonality of the target items. In this case, a choice of ‘1’ on the scale represented very uncommon, while a choice of ‘7’ represented very common. For example, participants may have seen a burst tyre, and were required to decide how common it would be to see this item in everyday life.

Procedure. The procedure for this experiment was identical to experiment 3.

Results

Figure 3.6 shows that, on average, participants responded faster in the prime condition (M = 981, SD = 269) than the control condition (M = 1067, SD = 324): *F*1(29) = -2.410, p = .023; *F*2(19) = 2.904, p = .009. As before, response times that were more than 3 standard deviations above the mean were replaced with the mean plus 3 standard deviations. Response times that were more than 3 standard deviations below the mean were replaced by the mean minus 3 standard deviations. Participants' responses were very accurate (M = 90%, SD = 4%), and there was no difference between the first half and second half of the experiment with respect to accuracy (p > .05). There was also no difference between conditions with respect to accuracy (p > .05).

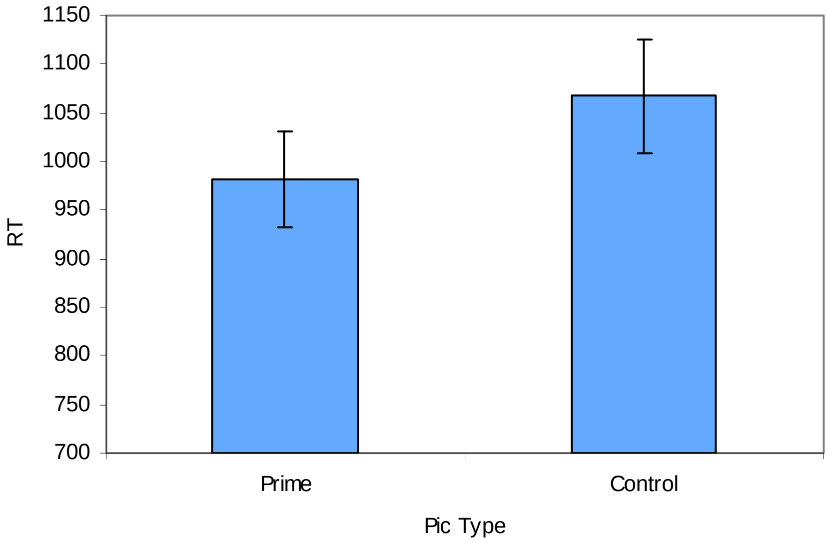


Fig. 3.6. *Mean participant response times for both conditions, with standard error bars.*

Norming. Table 3.5 shows the means and standard deviations for ratings for each of the three questionnaire versions. The mean figure for ‘Plaus’ represents the average rating out of 7 given to prime items for how likely a given event is to occur. The mean figure for ‘Likelihood’ represents the average rating score out of 7 given to items for how likely an event is to occur, given the instrument which precedes it. Lastly, the mean figure for ‘Commonality’ represents the average rating out of 7 given to targets in response to how common these items are deemed to be.

	Plaus (prime)	Likelihood	Commonality
Mean	4.87	5.45	3.32
SD	1.24	0.93	1.29

Table 3.5. *Means and standard deviations for each of the three norming types.*

Table 3.6 shows Pearson’s correlations between response times in the priming condition and each of the rating types. Negative correlations between response times for prime items and both the plausibility of events and the commonality of targets suggests that the faster the responses were, the more plausible the events were, and the more common the targets were. The significant correlation between event plausibility and likelihood indicates that the more plausible the events were deemed to be, the more likely it was deemed that the change in the target was brought about by the instrument.

	Plaus (prime)	Likelihood	Commonality
Prime RTs	-0.13*	0.06	-0.24*
Plaus (prime)	-------	0.72*	0.02
Likelihood	-------	-------	-0.14
Commonality	-------	-------	-------

*p<.05

Table 3.6. *Pearson's r values for correlations between knife-tyre prime and control condition response times and each rating type.*

Discussion

Participants in this experiment continued to respond significantly faster in the prime condition compared to the control condition. This indicates that these effects are not driven by participants' construction of associations between instruments and targets, but are instead driven by a combining of affordances to construct appropriate event-based representations. As before, the events coded in these representations contain information concerning the causal nature of the occurrences depicted by the pictures.

In this experiment, differences between conditions were significant in both by-participants and by-items analyses. The standard deviations of average participant response times were also smaller for this current experiment, indicating less variance in the time taken to respond to targets. Not only is this current set of items better controlled, but they are also more effective at bringing out the effects reported in these experiments.

As mentioned previously, although these experiments do not make use of a linguistic task, the possibility that participants mentally label presented items cannot be dismissed. However, it should be possible to bias participants' responses, to induce more automatic, rather than strategically controlled processing (Sachs et al., 2008) As combinatorial effects have been observed even when the stimulus exposure duration is as short as 250 ms, this is a potential way to ‘tighten’ the conditions under which these effects are borne out.

Experiment 5

Short SOAs (stimulus onset asynchrony) in priming paradigms (e.g. 100-250 ms) are thought to elicit fast, more implicit processing of the stimuli, while longer SOAs (e.g. 600-1000 ms) encourage more controlled, strategic processing (see Sachs et al., 2008; Neely, 1991). Sachs et al. (2008) suggest that although more strategic processing may reflect the organisational structure of the items being processed, less conscious processing may reflect the operation of automatic processes operating earlier on, before the items have been categorised or arranged thematically.

In fact Rossi et al. (2001), using related and unrelated primes, have demonstrated activation of different regions of anterior cingulate cortex, depending on the type of SOA used (either short, at 200 ms, or longer, at 1000 ms). This suggests that strategic or automatic processes are controlled by different brain areas, and can be induced with different paradigmatic constraints. Although combinatorial effects have been demonstrated in the current experiments under conditions generally considered to induce controlled processing, it is useful to establish whether these effects can also be induced under conditions thought to tap automatic processing.

Method

Participants. 40 University of York students participated in the experiment for half-an-hour's course credit or £2. All participants had normal or corrected-to-normal vision.

Materials. The same items and materials from experiment 4 were used in experiment 5.

Procedure. The procedure for this experiment was identical to that of experiments 3 and 4, except for different exposure durations of the first presented pictures. First pictures in this experiment were presented for 250 ms, instead of 2000 ms, and the duration of the second fixation cross was shortened to 100 ms.

Results

The slight difference between conditions (see figure 3.7) was only significant by-items (prime: M = 835, SD = 456, control: M = 882, SD = 456; *F*1(39) = -1.425, p = .162; *F*2(19) = 2.115, p = .048). The same trimming of response times to three standard deviations above or below the mean was used for this experiment. Again, participants' responses were very accurate (M = 96%, SD = 6%). There was no difference in accuracy from the first to the second half of the experiment (p > .05), and there was no difference in accuracy between conditions (p > .05).

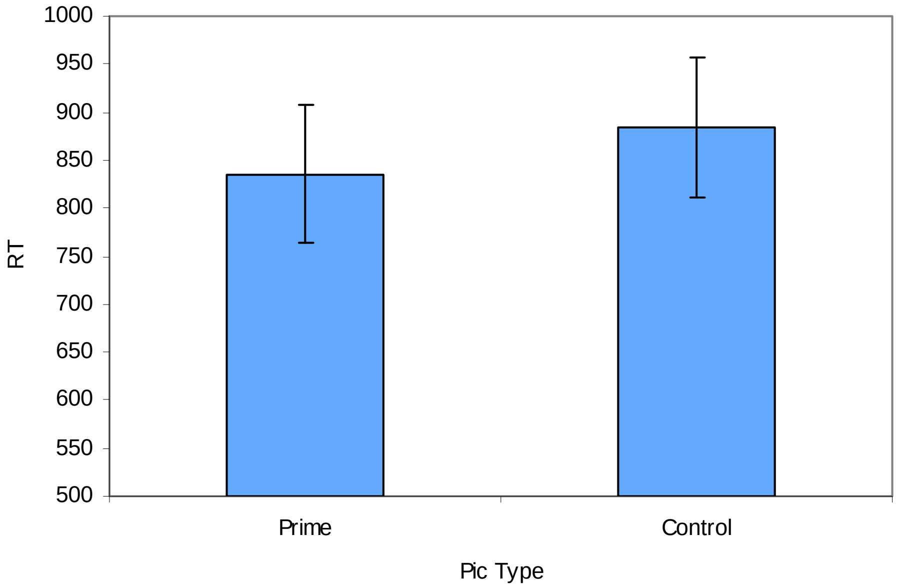


Fig. 3.7. *Mean participant response times for both conditions with first picture exposure durations of 250 ms, with standard error bars.*

Discussion

Across stimulus items, responses were significantly faster to the prime, compared to the control condition. Across participants, however, this difference was not consistent. The fact that the difference in average participant response times was not significant may be due to individual differences among participants (such as differing processing speeds). Alternatively, a larger variance in participant responses times may be due to experiential differences, given that affordances are derived on an experiential basis. In addition, participants' average response times in this experiment were significantly quicker compared to participant response times in the original 'knife and tyre' experiment

(experiment 3)[4]. This suggests that the shorter stimulus presentation times in this experiment were inducing participants to make their responses faster also.

Consonant with data from experiment 2 (lexical decision priming paradigm with 100 ms stimulus exposure durations), the combinatorial effects are considerably weaker when participants are not viewing the stimuli for as long as is necessary to construct event representations from item affordances.

General discussion

To summarise the results of experiment 3, for the bird-puddle items, although participants were significantly faster to respond in the control condition compared to the priming condition, the removal of outliers demonstrated that this effect was statistically fragile. For the knife-tyre items, participants were significantly faster to respond in the prime condition compared to the control condition.

This combinatorial effect observed for the knife-tyre items in experiment 3 was replicated in experiment 4, where participants continued to respond faster to primes compared to controls. In this experiment, stimuli were controlled for associations between instruments and targets, and the stimulus set was also slightly larger. This replication of the original effect emphasises the robust nature of the current paradigm and provides a rationale for its continued use.

Experiment 5 used essentially the same paradigm, but with shorter exposure durations (250 ms) of the first picture in each trial. Objects were

---

[4] Prime conditions: p = .02; control conditions: p = .01.

responded to faster in the prime condition than in the control condition, and although this was consistent across items, the by-participant analyses showed that participants were not consistently responding faster in the prime conditions than in the control conditions. This may have been due to individual differences among participants involving different processing speeds or respective experience with stimulus items. Participants in this experiment were also, on average, making faster responses than in the original ‘knife and tyre’ experiment, which may have diluted some of the differences between response times between the conditions.

Bird-puddle items

For these experiment 3 items, it was assumed that increased priming to a glass when participants were shown a bird and a puddle (compared to a boy and a puddle) would reflect participants’ recruitment of abstract representations, while the opposite would reflect the use of more concrete representations.

The fact that participants responded faster to the bird-puddle items in the control condition, although puzzling, suggests that participants are not recruiting abstract representations of events to aid object recognition. These data did not appear to be statistically reliable, however, given their sensitivity to the precise trimming criterion used. But still, if one is to assume that when participants see a bird and a puddle, a representation of the bird drinking is activated, the delay in reaction times may be due to a mismatch between this representation and the representation activated by the glass. Instead of complementing each other, however, these representations may compete, leading to longer response times.

With respect to when participants see a boy and a puddle, if one assumes that a representation is activated only with an appropriate combination of items, there should be no pre-activated representation of drinking from the puddle when it is paired with the boy. So, recognition of the glass in this case is free from any influence of event representations and so reaction times are faster on these trials, compared to trials where event representations (birds drinking from puddles) *are* activated.

Notwithstanding the difficulties inherent in interpreting a null result (or a result in a direction opposite to that which was predicted), the results from these items suggest that the activated representations are unlikely to contain general, abstract information pertaining loosely to the initial objects presented to the participants. If they did, we would have expected to see a reliable priming effect, with responses in the prime conditions faster than the control conditions. What is more likely is that the representations code the concrete interactions between the agent and the theme, and it is this information that is used to aid comprehension of the visual event. Participants viewing events depicted by the combination of bird-puddle items are unlikely to have extensive experiential knowledge of them and within the context of Colcombe and Wyer's (2002) account of event comprehension, participants are unlikely to draw upon abstract 'scripts' of events with which, although they have been observed before, the observer has not had direct experiential contact (such as a cow chewing grass).

A potential criticism of the bird-puddle half of experiment 3 is that a more effective paradigm might be to compare 'abstract' and 'concrete' conditions in a similar experimental design. The targets in the 'abstract' case would consist of objects that afford the event but are not associated with any interaction between agents and themes (e.g. a bird and puddle for the first picture, and a glass for the second), and targets in the 'concrete' case would consist of objects that appear as a result of some kind of interaction between the agent and theme (e.g. a bird and puddle for the first picture, and a smaller puddle for the second). Experiment 9 (chapter 6) utilises such items and expands this notion of concrete versus abstract representations.

Knife-tyre items

For the knife-tyre items in experiment 3, participants were significantly faster to respond in artifact judgements to targets in altered states when they were preceded by a combination of objects that afforded the state change, compared to when presented with control items that did not afford this alteration of the target. This effect was replicated in experiment 4.

Priming to target objects is assumed to arise from the combined affordances of the instruments and targets in the first presented pictures. Not only is the event itself primed, but the likely end-state of the object in the first picture is primed also. Priming is not directly attributed to affordance information, however, but from a generated event representation corresponding to the types of interaction that could occur between the instrument and object in the first picture. The causal state of the target object (e.g. a burst tyre) is coded within the representation that is constructed prior to any processing of the target object occurs.

Causal information is important for our perception of the external world and in part drives the formation of representations of events around us (White, 2006). Inferring causality from visual displays has been demonstrated extensively with the use of moving 2D displays. Michotte’s (1963) seminal work on causality and perception for example, suggested that a fast, stimulus-driven, automatic process operating on participants’ perception of moving blocks allowed them to infer a causal relationship between the two objects. Scholl and Tremoulet (2000) attribute this inference to the high-level perception of causality and animacy ‘making contact’ with lower-level processing of simple geometric shapes, devoid of any semantically relevant information.

Of course, the stimuli used in the current experiment do not provide such a visually impoverished signal, but are, in addition to causal information, bestowed with rich semantics. The causality implied by these current stimuli do not come from motion as in the Michotte studies, but arise instead from the combined affordances of the objects in the first presented picture. Crucially, however, in the current paradigm the affordances from items can only be combined with respect to the event they both afford. The representation is essentially of an event that entails something being altered or changed in some way by the object that is beside it in the scene. It is this in this way that causality is represented in the current stimulus presentations.

The findings from the knife-tyre items also hold implications for the study of object recognition. Participants in experiment 3 and 4 were constructing event representations on the basis of the knife and tyre (which together reflect the likelihood of an interaction between the two). Such an interaction most plausibly results in the tyre being burst, and this representation, constructed in response to the prime therefore facilitates recognition of the target.

This account of the data is based on the assumption that event representations are generated from combined affordances from the *first* picture. It is this representation that is then used to aid recognition of the target object: an object that represents the prior presented object in its end-state. However, there is another scenario which could be used to account for the knife-tyre data presented in the current chapter. This account presumes that the affordances of the instrument and object in the first picture are only integrated into an event representation upon seeing the target object in its end-state. This account assumes that no combining of affordances occurs upon seeing the first picture, but instead, that the target's end state is 'ratified' by the prior presentation of relevant objects, and it is this which induces facilitation of responses to the target. This notion of integration of information at the second picture is to be investigated in an experiment presented in the next chapter.

For now, however, the fact that combining objects to form an event representation impacts object recognition is perhaps not surprising. Zacks & Tversky (2001) make an interesting observation concerning object recognition in real-world circumstances. Usually, observers experience an exact event, happening at a specific time, only once, while experience with objects allows repeated viewings, re-examinations and a wide range of other experiential contact. These authors argue that our perceptions of events are driven by conceptual expectations and prior knowledge of what that particular event might entail. It seems beneficial, therefore, to make use of causal structure information in comprehending an event as we may not be able to exploit repeated, concentrated visual examinations of it to facilitate visual processing.

Representations containing causal properties are thus recruited in order to compensate for the likely temporally transient nature of the unfolding event.

Shortened exposure durations

In experiment 5, participants responded faster in the priming condition (e.g. a knife and tyre for the prime, and a burst tyre for the target), compared to the control condition (e.g. a knife and ruler for the control, and a burst tyre for the target). This difference was only significant in a by-items analysis (differences in response times averaged across items), however. A larger variance in participants' response times across conditions seems to have induced a slightly weaker effect and may have been due to differences in participants' respective experiential knowledge. Also, this more constrained paradigm has elicited faster responses in each condition compared to previous experiments, again a likely result of the reduced exposure duration of first pictures in each trial.

According to Neely (1991), stimuli presentations times of 250 ms or less is short enough to tap automatic processes, while longer durations (600-1000 ms) elicit strategic control over stimulus processing. These strategies include attempts to develop relations between every first and second picture (although relations can be established during prime trials, participants would not expect these relations on every trial because of the inclusion of control trials and a large number of fillers). Strategic methods of completing the task such as rule development were kept to a minimum in the current experiments due of the large number of filler items – a small number of primes compared to filler items is assumed to reduce demand characteristics (Brown & Besner, 2002; Neely, Keefe, & Ross, 1989). Also, it would be unlikely for strategic control to account solely for the results in experiment 5 as the combining of affordances still seem to be borne out in a more constrained paradigm, albeit with slightly weaker effects.

The experiments described in this chapter serve to expand upon the original findings concerning the combining of semantic information from two objects and its subsequent facilitation of object recognition. In addition, these experiments demonstrate the construction of representations containing event and causality-based information. These effects are also elicited with the use of modified time parameters, although to a lesser extent, betraying perhaps a more stressed object recognition process.

There are, however, two assumptions of these data that provide the basis for experiments to be described in the next chapter. Firstly, if priming is driven by the combining of affordances from both items in the first picture, then no priming should arise in cases where an instrument is shown on its own, and any facilitation must occur from the integration of affordances with the target *after* the target has been viewed. Secondly, if priming is driven by a ‘bursting tyre’ event, then no priming should arise in cases where the target is shown as unchanged (e.g. an intact tyre).

Chapter 4

In the previous chapter, three experiments were presented, using a novel paradigm which sought to constrain the conditions under which combinatorial effects could be elicited. For the first two experiments, participants were significantly faster to respond in an artifact judgment task to a target object in a changed state (e.g. a burst tyre) when they were presented with preceding objects whose combined affordances suggested a particular event (e.g. a knife and tyre suggesting bursting). The stimuli in these studies comprised visual depictions of physical objects which, coupled with the artifact judgment task, eradicated the possible influence of linguistic information. The previous paradigm was also modified to ensure no associations could be made between targets and prior presented objects (regardless of the instrument it was paired with), and in an attempt to distinguish between automatic and strategic processes, first picture presentation times in the third experiment were shortened from 2000, to 250 ms. However, in this last experiment, a difference between prime and control conditions was only observed in a by-items analysis.

Thus far, these results have been interpreted in terms of the construction of event representations from the combined affordances of the objects presented in the first display. These representations aid recognition of the target object in the second display, resulting in a facilitation of response times to the target. Discussion of previous experiments has also focused on the notion that it was not one sole item from the first display driving the priming we saw in these cases, but both items in that display together. However, an alternative account introduced in the previous chapter is that recognition of the target is not driven by combinatorial processes at the first picture, but is instead driven by integrative processes at the target.

Using the trial items for the concept ‘burst’ as an example, the integrative account would run as follows: The object and instrument in the first picture (i.e. the knife and the intact tyre) are processed and the affordances of these items stored in short-term memory. Crucially, however, these affordances are not combined to form a representation of the event that these items afford. Instead, upon viewing the burst tyre target, the representation of that target includes the fact that it is burst and that something must have caused it to be in this state (cf. the causality implicit in linguistic expressions such as “the boat sank” which naturally continue “…an iceberg ruptured its hull” or “it had been torpedoed” – in effect, the verb “sank” begs the question “why?”). The availability in memory of a potential instrument that caused the bursting affords the integration of the knife and the intact tyre (reflecting the initial state of the tyre) with the burst tyre (reflecting its final state). Thus, it is not so much that the burst tyre is recognised faster, but rather the fact of it being burst is understood faster.

Hence, according to this account, the representations of the original knife and tyre in the first display need not be combined into an event representation encoding the prior and successive state of the tyre – rather, the two objects are integrated through subsequently seeing the burst tyre, which affords their integration.

This account will be tested in the following experiment by presenting participants with some trials in which first pictures depict just a single object – corresponding to the instrument (knife) in the preceding studies, with no other object (e.g. the tyre) accompanying it. These conditions are to be contrasted with others in which first pictures contain instruments paired with alterable objects (i.e. as in the original studies; a knife paired with a tyre). If priming effects obtain in cases where first pictures contain a knife *and* a tyre, rather than a knife alone, the integration account would be considered untenable. Past results would, in this case be attributed to the combining of affordances to create an event representation that, in effect, anticipates the subsequent target object.

Targets are also to be manipulated in the following experiment. Continuing with the same example, if the knife and the tyre are activating a representation of a burst tyre based on their combined affordances, a facilitation of response times should not be observed for cases where the target is left in its intact state – unchanged by the instrument in the first picture. Thus, experimental targets in the following experiments are to be either in their end-state (as if altered by preceding instruments), or left unchanged in their intact state.

Experiment 6

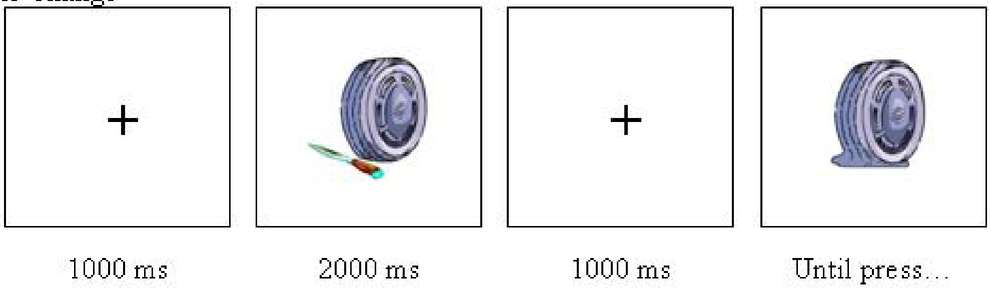
Figure 4.1 shows an example of the items used in the current experiment. The first condition contains the familiar priming condition used in previous experiments. In this condition, the knife, as before, is presented with the intact tyre, and the target is a depiction of this same object in its altered state. In the current experiment, any facilitation in this condition could be compared to a second condition in which the tyre is left unaltered by the instrument presented in the first picture. In this case, the event appears not to have occurred and so any event representation activated upon seeing the first picture cannot be exploited for the purposes of driving facilitation of response times to the target.

The third condition consists of, for the first picture, an instrument (the knife) divorced from its corresponding object (the tyre), with the 'missing' object in its altered state for the target. In the event presented for this condition, the instrument could only be interpreted as changing the target *after* the target itself has been viewed. According to Zacks and Tversky (2001), objects in semantic priming paradigms can be viewed by the observer either taxonomically or partonomically. The organisation of taxonomic information allows for reasoning concerning the properties of a particular object. For example, 'Ford' is a member of the category 'car' and we can infer that if a Ford is a car then it will contain wheels. Partonomic organisation is hierarchical, in that objects can be arranged by relations or parts and subparts. For example, observing legs enables one to infer standing (whether this be human legs or table legs).

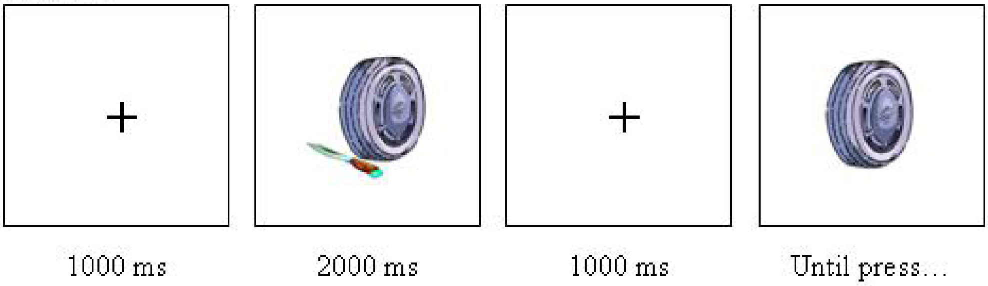
In the current experiment, a knife can be viewed taxonomically as a tool, and we can infer at the partonomic level that the presence of a sharp blade allows for a number of different interactions with it, including slashing, stabbing, cutting and so on. So, at stimulus exposure durations used in the current experiment, participants are able to glean a considerable amount of information from these instruments (whether they are presented on their own or not) upon their being presented.

In the last condition, the first picture depicts a knife without the corresponding object from the other conditions (e.g. the intact tyre), and the second picture depicts this object in its unaltered state. This condition essentially contains two constraints in that the instrument is not paired with its corresponding object, and the target is left unaltered. This condition can therefore be considered a baseline for the other conditions.

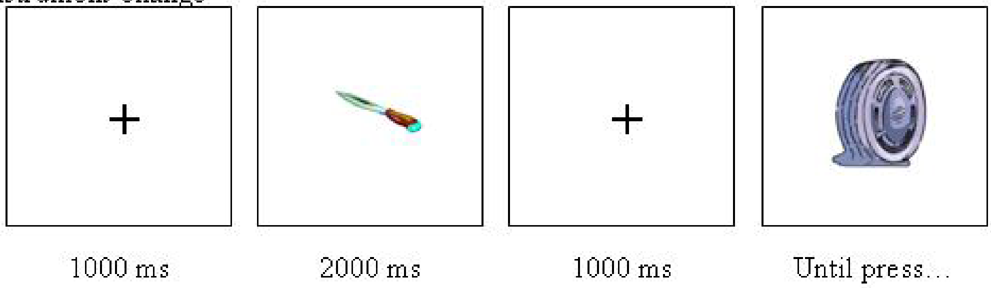
Prime-change



Prime-intact



Instrument-change



Instrument-intact

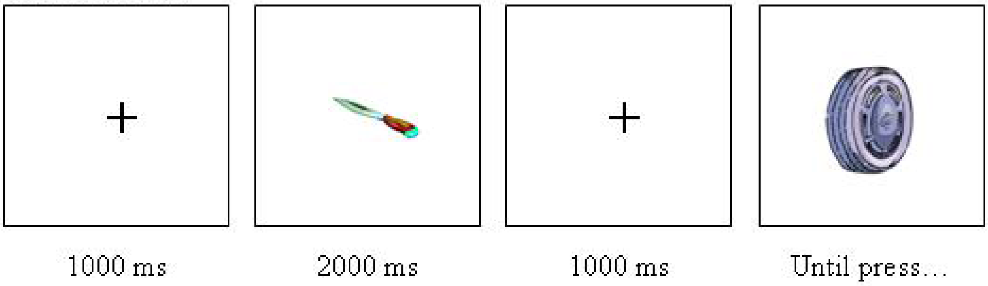


Fig. 4.1. *An example of stimuli used in each condition for the concept ‘burst’.*

Method

Participants. 60 University of York students participated in the experiment for half-an-hour's course credit or £2. All participants had normal or corrected-to-normal vision.

Materials. This experiment used the same 20 experimental items as experiments 4 and 5 (see appendix 5), and the fillers were altered slightly to take account of the fact that, for the first presented pictures, some of the experimental trials had an object and an instrument for the first picture (prime-change and prime-intact conditions), and others had just an instrument (instrument-change and instrument-intact conditions). As before, half of the trials required a 'manmade' response and half required a 'not manmade' response. There were four conditions, and each participant was presented with all 20 experimental items, five from each condition.

Each participant was assigned to one of four groups. Participants in each group would be presented with five experimental items from each condition, and these items were rotated across groups. For example, participants assigned to group one would be presented with the 'prime-change' stimuli for the concept 'burst', while participants assigned to group 2 might see the 'prime-intact' stimuli for this same concept.

Procedure. The procedure for this experiment was identical to that of experiments 3 and 4.

Results

Table 4.1 shows the means and standard deviations for each of the four conditions, and figure 4.2 shows that, on average, participants did not respond any faster as a function of the type of picture presented.

	Prime-change	Prime-intact	Instrument-change	Instrument-intact
M	998	938	971	981
SD	520	420	347	445

Table 4.1. *Means and standard deviations for each condition.*

A two-way repeated measures ANOVA was conducted, with one factor of picture 1 tyre (either: prime with knife *and* tyre, or knife alone) and a second factor of picture 2 type (either: altered targets such as a burst tyre, or unaltered targets such as an intact tyre). There were no significant main effects of picture 1 type ($F1(1, 59) = .080$, $p = .778$; $F2(1, 19) < .001$, $p = .997$), or picture 2 type ($F1(1, 59) = .913$, $p = .343$; $F2(1, 19) = .463$, $p = .504$). There was no interaction between picture 1 and picture 2 type ($F1(1, 59) = 1.459$, $p = .232$; $F2(1, 19) = 4.303$, $p = .052$) and none of the differences between conditions were significant (all $p > .05$). Average accuracy rates for participants were recorded (M = 90%, SD = 5%) and there was no difference between the first and second halves of the experiment with respect to the number of correct responses made ($p > .05$). There were no main effects of picture 1 or picture with respect to participants' accuracy rates, and no differences between conditions, with accuracy as the dependent variable (all $p > .05$).

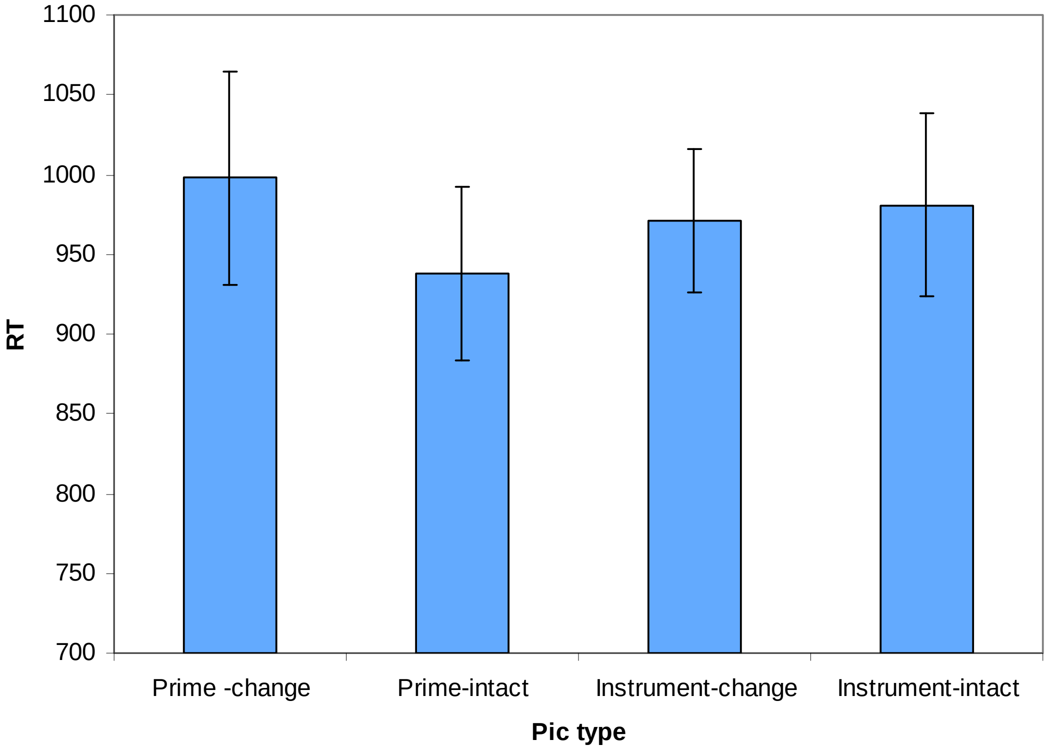


Fig. 4.2. *Average participant response times in each condition. Error bars represent standard error.*

Discussion

There were no significant differences between any of the conditions in experiment 6. Conditions in experiment 6 in which instruments were presented with no corresponding alterable object were predicated on their being able to test an alternative to the combinatorial account previously reported data. According to an integrative account of the data, priming effects from previous experiments are not attributed to the development of event representations from the combined affordances of two stimulus items. Rather, priming arises from an integration of relevant information *after* the target has been viewed.

Unfortunately, no conclusions can be drawn from the null results of the current experiment, with regard to this alternative integration account. The next experiment, however, will attempt to test this account again, with a simpler design, and with less of the design problems that are to be discussed below.

Firstly, this lack of priming observed in the current experiment may have been due to some general design problems which are to be discussed below. However, there is another perhaps more plausible explanation, and that is the differences between the numbers of items included in pictures between trials. For example, the prime-change and prime-intact conditions each had two items in the first pictures, while the instrument-change and instrument-intact conditions both had just one.

In cases where two items in one picture are presented to participants, it may simply be more demanding for the visual system to process two as opposed to one stimulus item. Studies of scene gist, for example, provide compelling evidence that scenes consisting of a large number items are more costly (in terms of the time-course of scene recognition) to process than a scene comprising just few items (Oliva, 2005). Object search paradigms have been shown to be sensitive to the amount of visual information contained within each stimulus presentation (Biederman et al., 1988), and differences in memory load due to the number of presented stimuli affects processing of the target (Sabb et al., 2007; Lavie et al., 2009).

Lastly, a design problem with the current experiment may also have contributed to the observed null results. As there were four conditions in the current experiment, but only 20 items, there may not have been enough power in the experimental design to reveal the combinatorial effects described so far. This issue is to be examined in the general discussion section of this chapter, but for now, the following experiment is to consist of two conditions, with the same 20 experimental items. This design has participants observing 10 experimental items and 10 control items in each experimental setting.

Experiment 7

Experiment 7 essentially splits up the preceding experiment into its ‘prime-change’ and ‘instrument-change’ conditions (see figure 4.1), and focuses on the notion of integrative processes at the target accounting for differences in response times, rather than combinatorial processes at the first picture. Figure 4.3 shows an example of the trial procedure for the concept burst, for both conditions. The experimental design of the current experiment is such that more power is available to reveal any effects. So, with just two conditions in this experiment, the changed instrument/intact instrument manipulation (‘prime-intact’ and instrument-intact’ conditions in the previous experiment) has been removed in favour of a more simplified design. However, it is important to mention that because experiment 7 effectively reuses two of the experiment 6 conditions in a simpler paradigm, it is still subject to the potential confound of unequal stimulus items across conditions.

As discussed previously, the processing demands may be much higher for trials in which two items are presented in the first picture instead of one. For example, regardless of whether significant effects reported previously were due to integrative processes upon seeing the target (as apposed to combinatorial process at the first picture), an imbalance between the numbers of stimulus items in first pictures may be contributing to processing speeds of the target. If we assume that information from the first picture in each trial is retained in short-term memory, then the memory load for one item in a scene should be much less than for two, and so the processing of the target may be more costly in the second, two item case, because more information is held in short term memory during the processing of target.

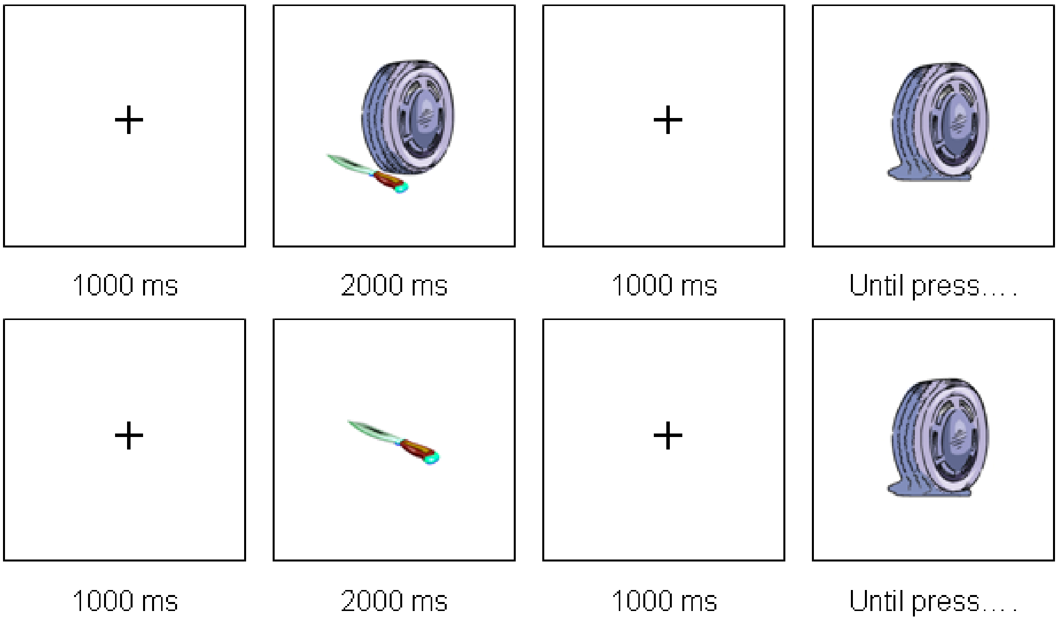


Fig. 4.3. *An example of the trial procedure in each condition for the concept 'burst'.*

Method

Participants. 30 University of York students participated in the experiment for half-an-hour's course credit or £2. All participants had normal or corrected-to-normal vision.

Materials. This experiment used the same 20 experimental items as experiments 4, 5 and 6 (see appendix 5). The fillers for this experiment were altered slightly so as to correspond to the experimental items. For example, some of the first picture filler items had only one item in the 'scene' to reflect the experimental cases where instruments were presented without a corresponding alterable object. There were 2 conditions, and each participant was presented with all 20 experimental items, 10 from each condition.

Participants were assigned either to group one or group two, and experimental items were rotated across groups. For example, a participant responding in group one might receive the 'prime' items for the concept 'burst' while group two participants might receive 'control' items for this concept.

Procedure. The procedure for this experiment was identical to that of the previous experiment.

Results

Figure 4.4 shows the average response times to targets for each condition. There was no difference between the two conditions (prime: M = 931, SD = 279; control: M = 933, SD = 299; *F*1(39)<1; *F*2(19)<1). Accuracy rates across participants were high (M = 96%, SD = 4%), and no improvement effects were observed from the first to the second half of the experiment (p > .05). In addition, there was no influence of condition on participants' accuracy, with no difference between conditions with a dependent variable of accuracy (p > .05).

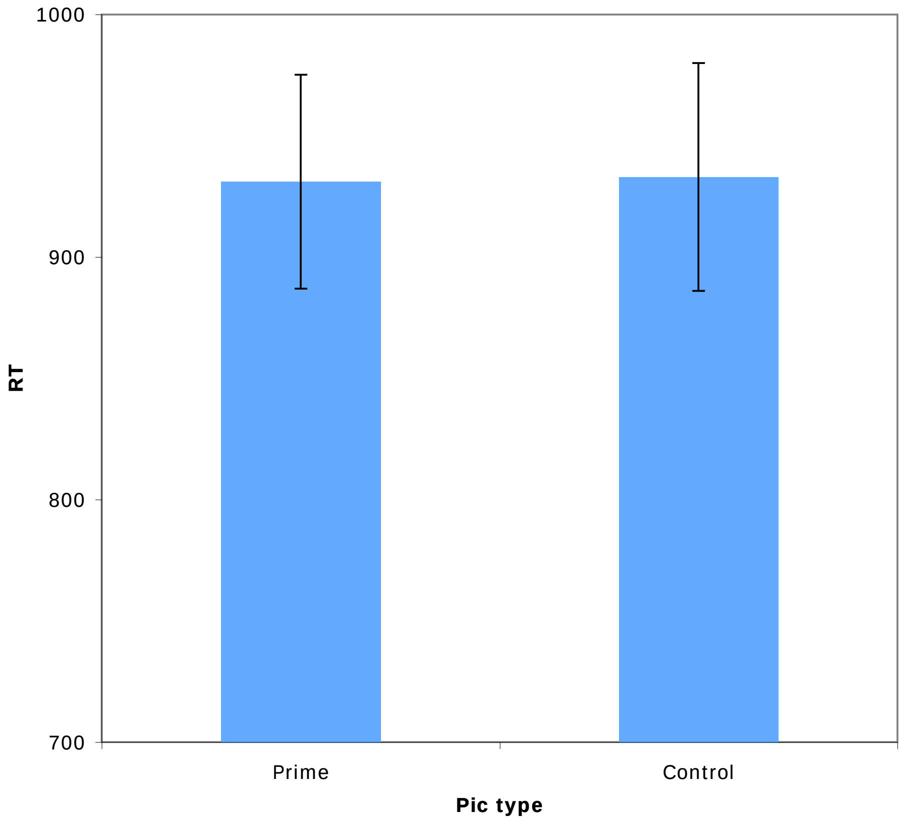


Fig. 4.4. *Average response times for both conditions.*

Discussion

Participants on average responded only slightly faster in the prime compared to the control condition, and this difference was not significant. If priming in previous experiments is being driven principally by the event representation constructed upon combining the affordances of previously seen items, it appears problematic that there is no apparent facilitation over a condition where such representations cannot be recruited.

The conditions in experiment 7 were used to test the assumption that priming in previous experiments was due, not to combinatorial processes at the first picture, but integrative processes at the target. We know that priming effects have been achieved using primes with an instrument and object which are conducive to a particular event (albeit with different controls to those used in the current experiment). So for the second condition in experiment 7 to elicit similar average response times is certainly not inconsistent with an integration account if one assumes that priming obtains in *both* conditions. However, as these data are null, this experiment is still inconclusive.

The next experiment aims to, again, establish how these priming effects are borne out, and also to address an issue concerning the lack of an adequate control condition, which is discussed below. Experiment 8 will consist of two conditions – a prime in which (for the concept ‘burst’) a sole knife would be presented, with a burst tyre for the target. And a control in which a sole ruler would be presented, with, again, a burst tyre for the target.

If, in the current experiment, priming effects were elicited form the sole knife condition, then average response times to these items would be expected to be faster than a control with a sole knife for the prime. With respect to the integration account, a burst tyre can be integrated with a knife (the knife provides a causal explanation for the state of the burst tyre), but a ruler cannot.

Another issue that is to be addressed in the next experiment is that the previously reported priming effects may have arisen in part from the control items. This is based on the assumption that facilitation of response times is not driven solely by priming from experimental items, but is also driven by *inhibition* from the controls. For instance, if participants are presented with a ruler and an intact tyre for the first picture, it is unlikely that any event representation could be generated from the combination of these items’ respective affordances, including that of a bursting event. However, upon seeing the burst tyre, participants' response times may be inhibited by a 'checking' of the target against information concerning the preceding picture held in short-term memory. The slower response times in this case arise from the additional time taken to match the target to the prior presented scene. Any such matching process can be thought of as mentally revisiting the previously seen items to ensure no event could have been possible, given their pairing.

Support for this explanation is based on the fact that both facilitation and inhibition are commonly implicated in semantic priming paradigms (Neely, 1991). As McNamara (2005) notes, facilitation and inhibition operating in semantic priming paradigms are typically inextricably linked and as yet, no paradigm has successfully teased the two apart.

Every condition used in experiments 6 and 7 contain first picture items which are able to interact with the target in some way. If we are to assume that average response times to the target are facilitated in each of the conditions in experiment 6 and experiment 7, a control condition containing an instrument which is *unable* to alter the target (and thus inhibit response times to the target) may be an imperative inclusion in subsequent paradigms for eliciting such effects as reported thus far.

As mentioned previously, the conditions used in experiment 7 contained an unequal number of items across conditions. In order to eliminate any potential confound of differing constraints on visual processing, the subsequent experiments shall ensure the amount of visual information remains constant between conditions.

Experiment 8

Experiment 8 reintroduces a control condition consisting of an object and an instrument which is *unable* to instigate a change in the target. Figure 4.5 shows an example of the trial procedure for the concept 'burst'. First pictures for the prime condition in this case consist of an instrument without its corresponding object, and for the control condition, an instrument (also with no corresponding object) that is infelicitous to a given event.

As this experiment has the same number of items in first pictures across conditions, and only two conditions, it does not suffer from a lack of experimental power (from the ratio of conditions to items) or the previously discussed confound concerning differing amounts of visual information across conditions. If priming from instruments alone is occurring in previous experiments, we would expect facilitation here to be evident in the prime condition, resulting in significantly faster response times to the target.

Method

Participants. 30 University of York students participated in the experiment for half-an-hour's course credit or £2. All participants had normal or corrected-to-normal vision.

Materials. Again, the same 20 experimental items as the previous four experiments were used, and the fillers were constructed so as to reflect these items. None of the filler trials contained two items in the first picture as none of the experimental items did either. Just as in previous experiments, participants were assigned to one of two groups and the prime or control trials were rotated between them. Again, participants responding in either condition received 10 prime trials and 10 control trials.

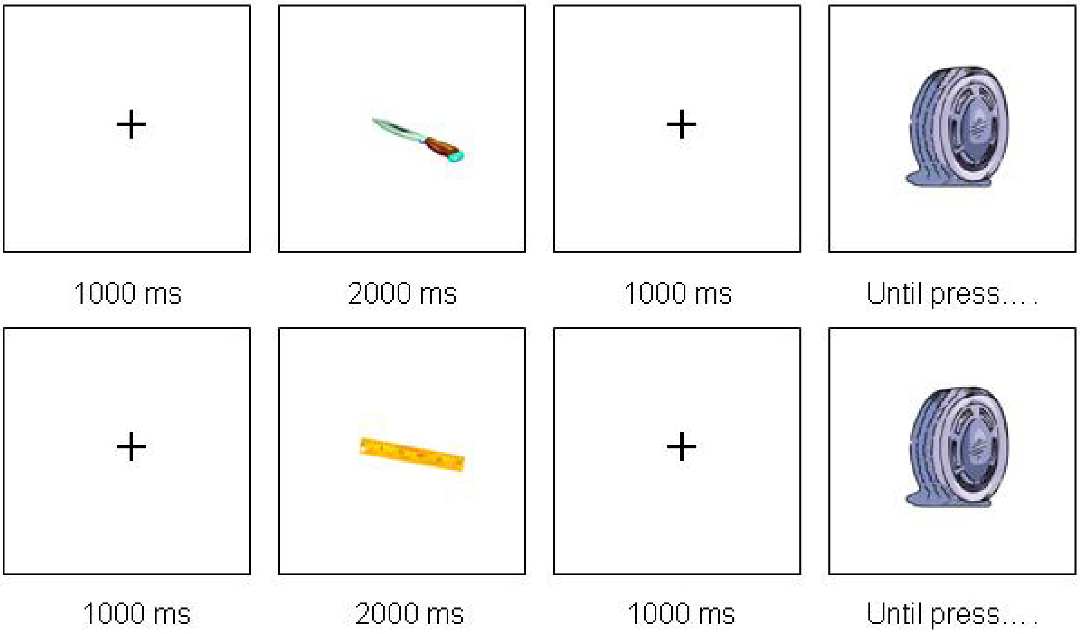


Fig. 4.5. An *example of trial procedure for the concept 'burst', for both conditions.*

Procedure. The procedure for this experiment was identical to that of the previous experiment.

Results

Figure 4.6 shows the average response times to targets for both conditions. There was no difference between prime and control conditions (prime: M = 1020, SD = 462; control: M = 1023, SD = 393; *F*1(29)<1; *F*2(19)<1). Participants responses were, on average, very accurate (M = 94%, SD = 5%) and no improvement effects, or a difference between conditions with respect to accuracy, were observed (all p > .05).

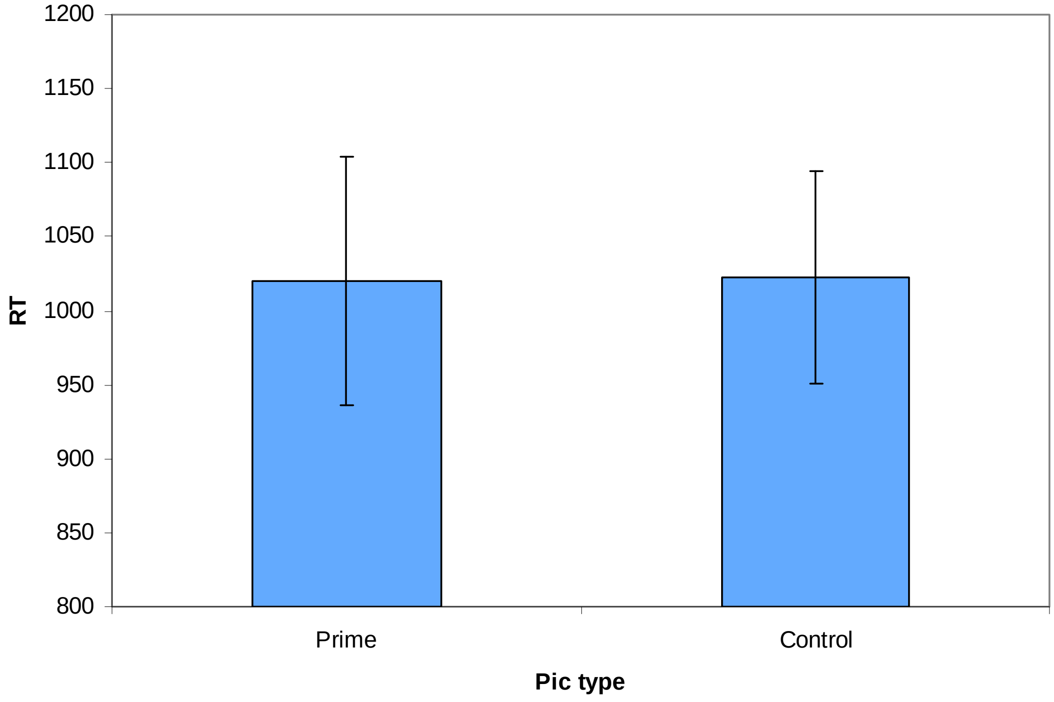


Fig. 4.6. *Average response times for both conditions.*

General discussion

The current chapter has presented three experiments, none of which have provided any statistically favourable results (at least in terms of the predicted outcomes). Each of these experiments will be discussed in this section in turn, with a summary of results for each, and a discussion of the data with reference to relevant literature.

Integration processes at the target?

In experiment 6, four conditions were presented to participants (see figure 4.1). First pictures in these conditions consisted either of objects and an instrument that could change that object in some way, or instruments with no corresponding object. Conditions where instruments were presented on their own were used to test an alternative to the combinatorial based account of previous data: That facilitation of response times is due to an integration of information both from the target and the preceding picture, but only after the target has been viewed. The targets for these conditions consisted of objects either in their end-states or still in their intact states. If causal properties were coded within event representations, priming should have only been observed in cases where targets were presented in their end-states.

The null results form experiment 6 make it difficult to draw any conclusions concerning the integration account, and it was unclear in this experiment whether there was enough power in the design to elicit the desired effects. As there were four conditions in the experiment, but only 20 items, this resulted in participants seeing only five experimental items from each condition. The small number of items relative to conditions is more likely to elicit Type II errors (Doan, 2005) and it is imperative this confound is avoided in the experiments that follow. For subsequent experiments, a two-condition design evades this problem, or if more conditions are necessary, additional stimulus items need to be constructed.

Another issue with experiment 6, which is also apparent in experiment 7, is that there was no appropriate baseline to which the other conditions could be compared. Semantic priming is a relative measure and so any facilitation is only apparent when compared to a neutral baseline. The ‘baseline’ conditions in experiments 6 and 7, however, were not neutral as two objects were presented to participants which could still plausibly interact with one another (e.g. a knife and an intact tyre). A more appropriate baseline would consist of entirely unrelated objects. In fact, several researchers have favoured the use of a simple array of x’s and other symbols in a word-based semantic priming paradigm for their baseline conditions (e.g. Neely, 1976; Posner & Snyder, 1975b).

Inhibition from controls?

In experiment 7, two conditions were presented to participants (see figure 4.3) in a simplified paradigm that aimed to test the integration account discussed previously. In the first condition, participants were first presented with an intact object and an instrument that is capable of altering that object in some way (e.g. a knife and an unburst tyre), and for targets, they were presented with objects in their altered state (e.g. a burst tyre). In the second condition, the same targets were shown to participants, but these were preceded by instruments on their own (such as a knife). If the facilitation of response times reported for previous experiments was due to combinatorial effects occurring at the first picture, rather than integrative processes occurring at the target, then participants should have responded significantly faster in cases where instruments were presented with objects in the first picture. This difference was expected because, according to the combinatorial account, participants are able to construct event representations to aid recognition of the target, based on the combined affordances of the instrument and object in the first picture.

There was, however, no significant difference between prime and control conditions in experiment 7. Assuming that priming obtained in *both* conditions (and hence no significant difference between them), the data for experiment 7 are consistent with an integration account of previously reported effects. If this account follows, the knife is just as effective as the knife *and* the tyre at facilitating recognition of the burst tyre. However, as these results were null and there may not have been enough items in these experiments to elicit the desired effects, these results are still inconclusive. While experiment7 sought to establish which account better explains the reported priming effects with a simplified paradigm, the difference between conditions in this experiment was also not significant.

A common feature of experiments 6 and 7 is that every condition contains an instrument which could plausibly interact with the target in some way. Perhaps previously reported effects were due, not solely to facilitation in the prime condition, but either to inhibition from the control, or both facilitation from the prime and inhibition from the control operating in parallel (Posner & Snyder, 1975b).

So, in experiment 7, it was unclear whether there was facilitation in either condition, or a lack of inhibition in the control condition. Neely (1976) suggests that a key factor in determining the respective contributions of facilitation and inhibition in semantic priming paradigms is the neutrality of the baseline and, as mentioned previously, the first pictures used for the control condition in experiment 7 were certainly not unrelated to the target. In fact, for a baseline to be considered purely neutral, it must be neither related nor unrelated to the target (Posner & Snyder, 1975b). Unfortunately, the use of such stimuli is untenable in the current paradigm, due to the nature of the inherent causal structure of the presented event. Also, a typical problem with purely neutral baselines is that they are not considered to have the same 'alerting' properties as other stimuli (McNamara, 2005), and so do not stimulate the visual system in the same way that a more relevant stimulus would.

Inhibition and facilitation are typically difficult to tease apart in semantic priming paradigms (McNamara, 2005; Neely, 1976), but facilitation is more likely to operate on response times at shorter SOAs. As the SOAs used in the experiments that contain more suitable controls (instruments which are unable to affect change in the target) are fairly long (3000 ms), it is likely that a degree of inhibition is operating on participants' responses in these cases. However, although weaker, a by-items difference in experiment 5 (first picture presentation times of 250 ms, a total SOA of 350 ms with the 100 ms fixation cross) was revealed at a shorter SOA. So it is still unclear as to whether inhibition is a sole contributing factor to the significant differences reported in previous chapters. What is most likely is that, consonant with Neely's (1976) findings, there is a degree of facilitation from the prime (that increases with shorter SOAs) and a degree of inhibition from the controls (that decreases at shorter SOAs).

Memory load?

Experiment 8 had participants responding to a prime condition in which the instrument that affords altering the target was shown on its own, with a target in its end-state, and a control condition in which an instrument which is unable to affect change in the target is shown, with an altered object for the target. For example, on a priming trial, participants may have been presented with a knife for the first picture, and a burst tyre for the target. And for the control condition, the first picture may have been a ruler, with a burst tyre presented for the target (see figure 4.5 for a schematic example of the trial procedure).

In experiment 8, visual information was controlled across conditions, an adequate control was used, and none of the items were repeated within trials. If inhibition from controls is a contributory factor in the facilitation of response times in this paradigm, then significantly faster response times should have been observed for the control condition in this case. However, there was no significant difference between the two conditions.

An interesting implication of experiment 8 was that, with regard to the integration account discussed previously, these data are inconsistent with the data reported for experiment 7. If it is to be assumed that in experiment 7, integrative processes at the target are responsible for priming obtaining in both conditions, then in experiment 8 we would expect responses to a prime with a sole knife to be significantly aster than responses to a sole ruler. So, while experiment 7 is consistent with an integration account of the data (albeit with a null result), experiment 8 is inconsistent with the same account.

An additional issue concerning the experiments presented in this chapter concerns the amount of visual information presented to participants across trials. As facilitation in the original ‘knife and tyre’ experiment has been replicated twice (albeit with weaker effects at shorter SOAs) it may be the case that other comparison conditions are simply given an unfair ‘advantage’.

In experiment 6 for example, of the conditions compared to the prime condition, two had less visual information for viewers to process, and one was subject to problems concerning the repeating of stimuli. Priming effects arising from presenting the same stimulus item to participants more than once, referred to as 'identity priming' (see Bruce & Valentine, 1985; Zipse et al., 2006), is thought to elicit stronger priming effects than those obtained through associations or links between items. It is, unfortunately, unclear whether the results reported in this chapter were contaminated by identity priming, given that there were no reported significant differences between any conditions in any experiment. Nonetheless, the notion of identity priming is related to the data presented in the next chapter, and is to be discussed further, below. With respect to the amount of visual information presented to participants in experiment 7, the comparison condition for the prime contained only one item in the first picture, and so again contained less visual information for them to process. So, if we assume that a knife and intact tyre / burst tyre condition (the 'original' prime condition) facilitates participants' responses, it may be that any perceived facilitation from instruments on their own is due to the amount of visual information held in short-term memory rather than the activation of any facilitatory event representations.

The role of memory load in semantic priming paradigms is well established (Beer & Diehl, 2001; Belke, 2008; Sabb et al., 2007), and responses to targets are often facilitated in tasks requiring less visual information to be held in working memory, compared to tasks in which the processing demands increase load on working memory. Sabb et al. (2007), for example, contend that processes involved with semantic priming are constrained by a limited-capacity system, and that an increase in working memory load (dictated by a larger number of stimulus items on ‘high load’ trials) decreases behavioural semantic priming. The assertion behind these authors’ work is that, due to the sharing of resources between working memory and semantic processing systems, this more efficient processing leads to faster behavioural responses if fewer stimulus items or features need to be processed.

Differing amounts of memory load from differing numbers of items across conditions may also have ensured its influence in the current paradigm because of the nature of the instructions given to participants: They were asked to attend carefully to every picture presented to them as there would be a memory test once the first phase of the experiment was complete. Although this instruction was used to avoid participants simply ignoring the first presented pictures (their task was to make decisions only on second pictures, after all), it may have strengthened any influence of differing visual information and memory load in the current experiments.

In experiment 8, visual information *was* controlled across conditions, but the condition with the knife presented on its own could not benefit from *comparatively* easier processing compared to the condition with the ruler. With regard to the experiments presented in the current chapter, faster responses in cases where an instrument is shown on its own may be due to a visual information confound, except for conditions which contain repeated items and so are subject to identity priming. According to this view, increased facilitation from identity priming is more beneficial to the processing of the target compared to any benefit gained from a reduction in visual information.

To summarise, no significant differences between any conditions were observed for experiments 6, 7 and 8. While certainly lending a degree of ambiguity to the results presented in previous chapters, there are a number of inherent problems unique to these current experiments which may help explain their findings. One of the properties which perhaps lends the largest amount of uncertainty to previously presented data is that of the relative contributions of inhibition and facilitation in semantic priming paradigms. However, these two factors seem to be modulated only by SOA, are difficult otherwise to separate, and still do not provide any edifying discernments concerning causality, affordances, and the mental representation of events.

Some of the current experiments were also subject to problems arising from a small number of items relative to the number of conditions, and identity priming of the object in the first picture to the target. While experimental power can be guarded against by employing a more carefully controlled experimental design, the notion of identity priming provides an interesting implication: if identity priming occurs only when the repeated stimulus is exactly the same as the prime (Neely, 1976), this suggests that the representation constructed upon initial processing of this stimulus is specific to the item seen. The nature of event representations with respect to their degree of ‘concreteness’ was examined initially in experiment 3 (see Chapter 3). For this experiment, participants were required to make an artifact judgement for targets which suggest the event depicted in the first picture, but only in an abstract sense. For example, participants may have been presented with a bird and a puddle for the first picture (depicting the event ‘drink’) and were required to make their decisions on a glass (*for* drinking, but not the type of drinking expected of the agent in the first picture).

If identity priming is an impacting factor in the current experiments, it may be indicative of something central to the nature of the semantic priming elicited by these experiments. This notion is concurrent with the idea that the representation activated upon seeing, for example, a knife and a tyre, is a representation of bursting which is specific to these items. If this is indeed the case, then we should expect no priming in a similar paradigm which contains, for the target, a tyre as before, but one which is clearly different from the tyre presented in the first picture. The experiments in the next chapter seek to examine this idea, and further elucidate the nature of the event representations activated in the paradigms described thus far.

The problems with the current set of experiments shall be avoided in subsequent paradigms by ensuring that, firstly, the number of stimuli presented to participants in each group for each condition is large enough to elicit the intended effects; that secondly, the number of stimuli is controlled between conditions; thirdly, that items are not repeated on given trials; and lastly, that control items are infelicitous to a given event (i.e. unable to affect change in the target).

Chapter 5

Introduction

In this chapter, a single experiment is presented which seeks to expand upon findings reported previously concerning causal structure in event sequences, the combining of affordances from objects in visual scenes, and the generation of event-based representations from these combined affordances. These previous findings suggest that observers of visual scenes are sensitive to the affordances of constituent objects, to the extent that a felicitous combination of affordances from an instrument and an object (e.g. a knife and a tyre) prime the likely change in state of the object (e.g. a burst tyre). An infelicitous combination of affordances from an instrument and object (e.g. a knife and a ruler), however, does not prime the same change in state (e.g. a burst tyre). The ruler in this infelicitous combination does not afford bursting the tyre and so that particular event is not primed.

It is assumed that this priming is due to the generation of event-based representations congruent with the types of affordance inherent in the presented stimuli. After their construction, these representations are recruited to aid recognition of the changed-state targets. For example, if the observer is presented with a wall and a sledge hammer, the affordances of the items (the wall affords having a hole knocked in it and the sledge hammer affords knocking holes in things) are combined to form an event representation of the hammer knocking a hole in the wall. With this representation still active upon processing of the target, the target's recognition is then facilitated.

In addition, these effects are unlikely to arise from integrative processes occurring when the target is viewed. This alternative account, discussed at length in the previous chapter, assumes that once a target such as a burst tyre has been recognised, the cause of its alteration is easier to ascertain if a prior presented context matches it (such as with a knife and an intact tyre), compared to when it does not (such as with a ruler and an intact tyre). This integration of information may be instigated by the fact that the target is in its non-canonical state, thus requiring a causal explanation. Although an instrument on its own, such as a knife, can be considered the most direct causal explanation for a burst tyre (as apposed to a knife with an intact tyre) there was no observed difference between a condition containing a sole instrument for the first picture, and a condition containing an instrument with an intact object.

In an attempt to expand the implications of the current results reported thus far, the experiment presented in this chapter aims to support the assumption that activated representations in the current paradigms are object-specific. In other words, the event representation activated for a pair of objects would contain information specific to the items seen, rather than information pertaining to the category of item, or a more abstract instantiation of a depicted event.

Representational codes

An obvious omission in the above combinatorial account of the data, and indeed, in much of the work focussing on representational accounts of behavioural findings (Block, 1983), is an indication of the way in which the representation is coded from the presented items. Are the affordances from instruments and objects derived from concurrent items directly, or from those items' respective categories? The category 'knife' might include such affordances as 'affords gripping', 'affords bursting', 'affords throwing', and so on. Similarly, the category 'tyre' might include affordances such as 'affords rolling' or 'affords bursting'. However, there may be specific affordances associated with a particular example of an object, independent of its category. For example, the category 'knife' might prime 'bursting', but what happens when the particular example of a knife observed has a blunt blade? Thus, it is important to establish whether the representations activated from instruments and objects are specific to these items, independent of the category to which they belong.

If we assume that targets in the current set of experiments are primed based on the combined affordances of the prior presented instrument and object, then object-specific representations would only manifest with 'tokens' of the same 'type' for the target, but not for different tokens of the same type. If we use the example of a knife and an intact tyre for the first presented picture, the intact tyre can be the same tyre as the target (albeit in its unaltered state). Or, it could be a different token of the same tyre – a tyre which, although still affords being burst by the knife, looks different from the intact tyre and is clearly not the same object.

Figure 5.1 shows an example of the trial procedure for experiment 9. The first two conditions consist of a prime and a control, with 'same-token' targets (the targets in this example are the same tyres as the intact version presented with its instrument in the first picture). The last two conditions consist, again, of a prime and a control, but with 'different-token' targets (the targets for these conditions are slightly different to the object presented in the first picture).

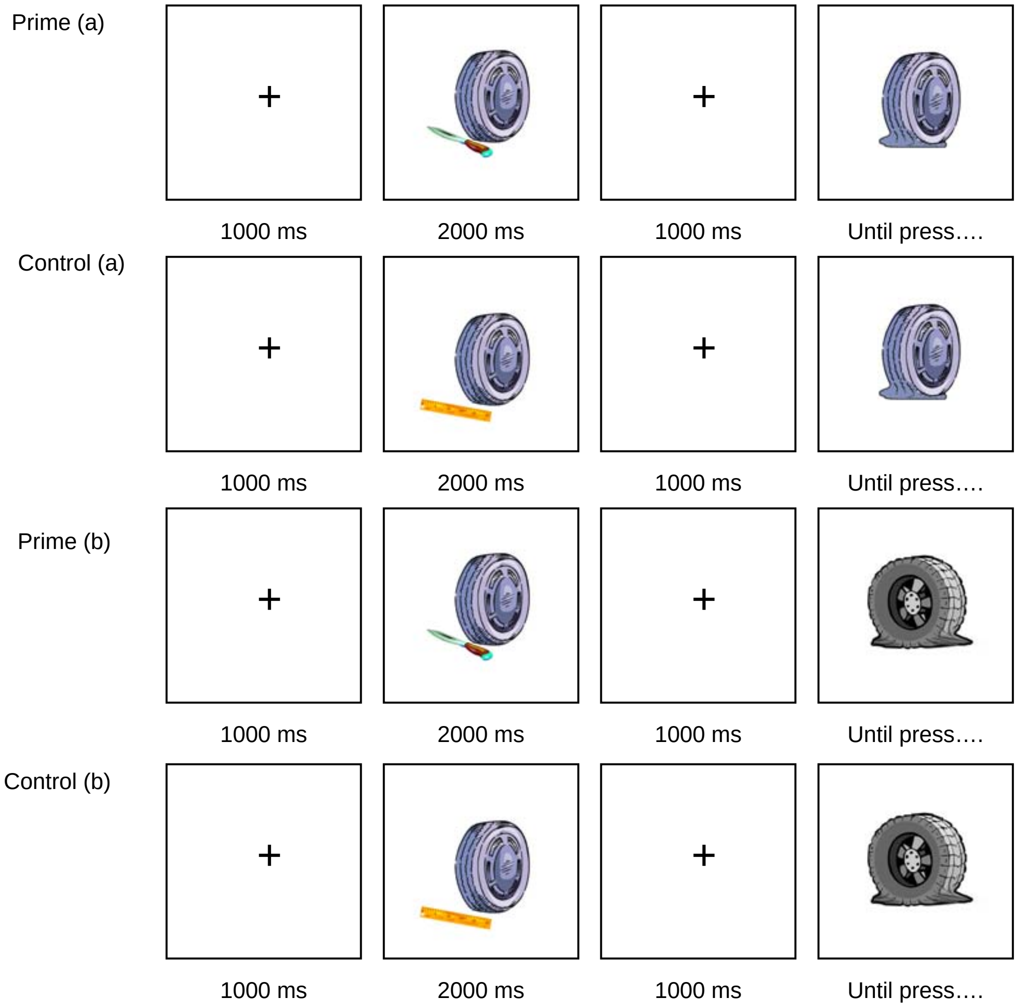


Fig. 5.1. *Example of the trial procedure for experiment 9.*

Maintaining the assumption that event representations activated in the current paradigms are object-specific, differences between prime and control conditions are expected, but only when same-token rather than different-token targets are presented. So, facilitation of response times to the target is only expected in the 'prime a' condition, as this is the only case where the representation generated from the affordances of the instrument and object can be used to aid recognition of the target. In the 'control a' and 'control b' conditions, appropriate event representations cannot be constructed from first picture items, and in the 'prime b' condition, recognition of different-token targets cannot be aided by representations consisting of different objects in intact states.

Perceptual grouping and the graspability of objects

An indication as to whether event representations contain low-level object-specific information independent of their high-level 'global' properties can be attained from studies of the functional grouping and 'graspability' of objects, and their influence on visual processing. Green and Hummel (2006) found that objects arranged so as to interact with one another were grouped perceptually, based on their mutually supportive function. Participants were required to make a decision as quickly and accurately as they could, as to whether the second picture in a two-picture sequence (e.g. a glass) matched a label given at the start of each trial. However, on some trials the first picture (e.g. a jug) was arranged so as to interact with the glass (e.g. arranged to afford pouring) and on others, it was arranged so as to not afford the event (e.g. positioned with the spout in the opposite direction to the rim of the glass). These authors found that participants were significantly faster to make their decisions on trials where the objects were arranged functionally, compared to when they were not.

Crucially, the non-interacting jug in this example *still* affords pouring. With regard to the interacting jug, however, it is only in its combination with an appropriately oriented glass that it affords pouring over and above what it affords in terms of its category. Its affordances are now specific to the object itself and its functional arrangement with other sequentially presented objects. These results suggest that, if the specific arrangement of objects is important in deriving these effects, then the specific object token may important also. In Green and Hummel's experiment, the event representation generated upon seeing the two pictures includes object-specific, or 'token' information, as apposed to solely general or 'type' information (see Kanwisher, 1987; 1991, Gajewski & Henderson, 2005, Chun, 1997).

Other studies have produced similar results with objects to which very little semantic knowledge is attached. Symes, Ellis and Tucker (2007) found that when objects to which no semantic 'object actions' could be attributed (such as a cylindrical rod) were oriented with edges readily 'graspable', participants were faster to make feature-based decisions relevant to the presented objects. Symes et al. attributed these effects to what they term 'pure physical affordances' (PPAs): Properties of visual objects which serve to afford action, independently of any semantic association. Observers in these experiments process what the concurrent object affords in terms of relevant actions, as apposed to processing what this particular class or category of object affords. These experiments suggest that, when processing a particular item, information pertaining to the object's potential for future interaction tends to concern token-specific, rather than token-type information. This type-token distinction is a fundamental tenet of the experiments presented in the current chapter, and forms the basis of the discussion concerning these experiments.

Object-specific representations and the visual world

The notion of object-specific event-based representations, however, seems counterintuitive to accounts of change blindness and incomplete representational codes of our visual surroundings (Simons and Levin, 1997, Rensink, 2000), and efficient or 'cheap' visual processing (Clark, 1999). Perhaps the most persuasive evidence that we do not build up rich, detailed representations of our visual surroundings comes from studies of change blindness. Relevant to the type-token distinction, small token changes to objects within a visual scene often go unnoticed if changes are made within a saccade. Mandler and Parker (1976), for instance, found that observers of a visual scene would not readily detect token changes to objects within a scene (for example, a mug changing to a different mug), but they would detect object-type changes (such as the mug changing to a plate).

It is likely, however, that in the observation of a scene encompassing multiple objects, arranged in front of a background, one typically makes several fixations in order to process the required information in the scene (Irwin et al., 1995). Very little detailed visual information is maintained from fixation to fixation, however (Henderson, 1997), and representations are not typically accumulated, but formed 'when needed' (White, 2006). Moreover, the inclusion of a background has shown to have facilitatory or inhibitory (depending on the contextual appropriateness of the background) effects on object recognition (Boyce & Pollatsek, 1992, Eriksson, 1989).

In the case of the current experiments, the presented scenes are simplified, do not contain backgrounds, and likely do not require as many fixations to process all the relevant information. In a complex scene, if allowed enough time, the eyes will make several fixations to objects, but representations from multiple objects will not accrue due to the processing constraints of visual short term memory (vSTM). According to Irwin (1991, 1996), the capacity of vSTM is only about three to five unique objects. Similarly, information specific to the mug in Mandler and Parker’s (1976) experiment is likely diluted by subsequent fixations to other objects, constrained by the capacity of vSTM. This dilution, thought to be analogous to visual masking (Campbell & Wurtz, 1978), is a common feature of visual information processing during saccades (Irwin et al., 1995; Irwin, 1991).

In the current experiments, the event representations constructed from these simple scenes are likely to be specific to the objects in the scene, including any specific features of those objects (shape, colour, size, and so on). In the absence of contextual information such as a background, observers tend to focus more on component features of the objects (Biederman, 1987). Information pertaining to context in this case can be considered the way in which these objects are able to interact with each other, given their combined affordances (see Green & Hummel, 2006). An explanation for why information is typically not accumulated from different objects within a scene was offered by Irwin (1991). It may simply be that there is no internal visual buffer with which to store these accrued pieces of information from successive fixations. Such a buffer may be obsolete in the current experiments, however, due to the small (at least in comparison to Irwin’s stimuli) number of items in each presented scene.

Current experiment

The experiment presented in the current chapter is a test of the assumption that the representations activated in these paradigms are based on token-specific codes, rather than more abstract, token-independent codes. The first and second conditions will contain the prime (e.g. a knife and tyre) and control (e.g. a ruler and tyre), and the target for these conditions will be an end-state object (e.g. a burst tyre). The third and fourth conditions will contain the same items but the targets shall consist of different tokens of the same type of end-state objects (see figure 5.1 for an example of the trial procedure). For example, the target for conditions 3 and 4 might contain a burst tyre, but with a token change (e.g. a burst, but slightly different tyre to the one presented initially with the knife).

As there are four conditions in the current experiment, to use the same stimuli as previous experiments would mean that participants only receive 5 experimental items for each condition. To increase the power of the experimental design and avoid committing a type II error, the stimulus set has been expanded from 20 to 40 items. So, participants in the current experiment will receive 10 experimental items from each condition.

Experiment 9

Method

Participants. 60 University of York students participated in the experiment for half-an-hour's course credit or £2. All participants had normal or corrected-to-normal vision.

Materials. Participants were assigned to one of four different groups. Each group received all 40 experimental items, 10 from each condition (see figure 5.1). Items were rotated across lists using a Latin Square, so participants assigned to group 1 might see the ‘burst’ concept in the ‘prime a’ condition, while participants in group 2 would see this concept in the ‘prime b’ condition, and so on. Filler trials were designed to correspond to the experimental items: first pictures consisted of two items. As before, half of the trials required ‘man-made’ responses, and half required ‘not man-made’ responses. All the experimental trials required ‘man-made’ responses. There were 160 trials in total, and the experiment lasted roughly 30 minutes.

Procedure. The procedure for this experiment was identical to that of the previous experiment (Experiment 8, chapter 4).

Results

Figure 5.2 shows that, on average, participants responded the fastest in the prime condition, compared to other conditions. A two-way repeated measures ANOVA was conducted with one factor of picture type (either prime or control) and a second factor of target type (either the same or different). Main effects of picture type (*F*(1, 59) = 4.860, p = .031; *F*2(1, 39) = 4.060, p = 0.49) and target type (*F*(1, 59) = 4.201, p = .045; F2(1, 39) = 3.448, p = .071) were revealed, but there was no interaction between picture type and target type (*F*1(1, 59) = 1.135, p = .291; *F*2(1, 39) = .175, p = .678). Planned comparisons revealed significant differences between ‘prime a’ and ‘control a’ conditions (prime condition with same-token targets compared to control condition with same-token targets (*F*1(1, 59) = 4.668, p = .035; *F*2(1, 39) = 4.211, p = .047)); ‘prime a’ and ‘prime b’ conditions (prime condition with same-token targets compared to prime condition with different-token targets (*F*1(1, 59) = 4.690, p = .034; *F*2(1, 39) = 2.629, p = .113)); and 'prime a' and 'control b' conditions (prime condition with same-token targets compared to control condition with different-token targets (*F*1(1, 59) = 7.496, p = .008; *F*2(1, 39) = 4.302, p = .045)).

Accuracy rates remained high for this experiment (M = 96%, SD = 4%), and there was no improvement from the first to the second half of the experiment (p > .05). In order to asses whether accuracy rates differed between condition, a two-way repeated measures ANOVA was conducted with one factor of picture type (either prime or control) and another factor of target type (either same or different), with a dependent variable of accuracy (number of correct responses across conditions). There were no main effects of picture type (*F*1(1, 59) = .280, p = .599; *F*2(1, 39) = .335, p = .566) or target type (*F*1(1,59) = .032, p = .858; *F*2(1, 39) = .089, p = .767), and there was no interaction between the two (*F*1(1, 59) = .301, p = .585; *F*2(1, 39) = 1.068, p = .308). Paired comparisons revealed no significant differences between conditions with respect to accuracy (all p > .05). Adjustments were made for all multiple comparisons using Bonferroni corrections.

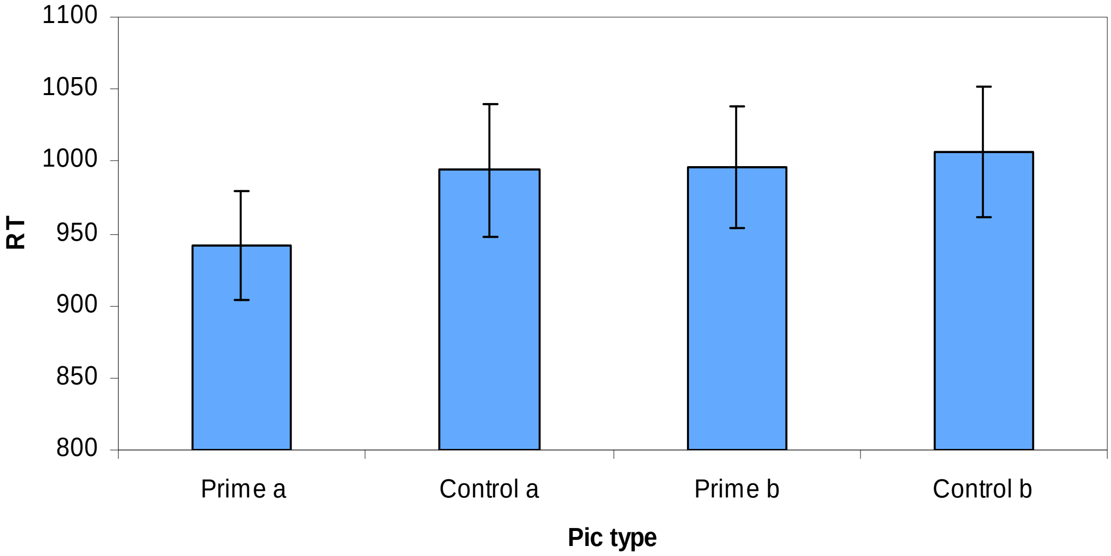


Fig. 5.2. *Average response times for each condition, with standard error bars.*

Discussion

Participants were significantly faster to respond in an artifact judgement task to targets in the prime, compared to the control condition, but only when targets were same–token objects. For example, responses to a burst tyre were facilitated when this same tyre was shown in its intact state beside a knife, but this facilitation was not evident for responses to different-token targets. In addition, response times were significantly faster in the 'prime a' condition (e.g. knife and intact tyre with same-token targets), compared to every other condition. These findings suggest that representations activated upon seeing the primes are specific to the particular objects shown. With respect to the accuracy data, participants' accuracy was no more or less likely to be higher depending on the condition, and participants did not appear to improve from the start to the end of the experiment.

Causal structure and scene comprehension. These data carry implications for the study of visual scene comprehension, object recognition, and the types of representation generated and utilised for the processing of visual scenes. Typically, the representations constructed upon viewing a visual scene are said to contain very little specific perceptual details, and it is only the basic semantics inherent in the constituent objects which are encoded in these representations (Grimes, 1996). Even when task demands encourage observers of visual scenes to form representations based on interactions or events between two objects, the information retained from the scene is considered to be sparse, with only simple spatial, and object-type information being encoded in these cases (Gilchrist, Humphreys & Riddoch, 1996).

One implication of these previous findings is that the observers of these scenes are forming general, abstract representations of them, much like a schema or 'script' of an event (see Ferretti et al., 2001; Rumelhart, 1980; Shank & Abelson, 1977), containing object-type rather than token-type codes. The current data refute this notion of representational coding, however, and demonstrate the coding of token information within event-based representations.

Possible mitigating factors in this interpretation of the current data are that of causality and object state changes. The current experiment essentially presents two types of event. In same-token cases the causal structure of the event is kept constant, with targets representing the 'end-state' of the depicted event. Instruments in this case are attributed with precipitating the change in state of the target object. In different-token cases, however, the causal structure of the event is disrupted, leaving the generated event representation unable to operate on target recognition.

This disruption of causal structure is discussed by Hommel (1998), within a framework encompassing the notions of episodic binding and code confusion. Hommel suggested that repeating certain features in a visually presented event binds them together in a single episodic representation (see also Kahneman et al., 1992 and van Dam & Hommel, 2010). This episodically-bound representation is automatically retrieved if one or more of the bound features are viewed again. Importantly however, if the repeated stimulus does not entirely match the features of the episodically-bound representation, then confusion between the two competing codes occurs, resulting in a processing cost for the presently viewed stimulus.

With respect to the current experiment, code confusion may occur between the slightly different target and the prior presented object. When participants were viewing a knife and an intact tyre, for example, code confusion may operate when a burst tyre is presented for the target when its features do not completely match that of the prior seen tyre. This notion of code confusion is also relevant to results presented in the previous chapter in which inhibition from control conditions was discussed as a contributory factor to the observed differences between prime and control conditions. According to Hommel's (1998) account, inhibition from controls in previous experiments would arise when, for the first picture, an object is paired with an instrument which is unable to change it (such as a ruler with an intact tyre) and, for the target, that same object in its altered state (such as a burst tyre).

Experiment 3 (see chapter 3) broached this issue of violations of the causal structure of events. In this experiment, participants were required to make an artifact judgement on a target (e.g. a glass) which represented an abstract instantiation of an event suggested by an agent and a theme (e.g. a bird and a puddle). As with the current experiment, less facilitation to the targets was observed in experiment 3 because the causal structure of the events had been violated. Using the above example, a representation of a drinking event is assumed to form upon seeing a picture of a bird and a puddle. This representation, however, is unlikely to encode a general drinking event (encompassing different forms of drinking), but instead appears to contain information specific to the type of drinking that could occur, given a bird and puddle. The concrete nature of this representation was such that it could not be utilised for the processing of the target (the glass). Furthermore, in order for causal event structure to be disrupted in the first place, token-specific information would have to be retained from the initial activated representation.

Distal and proximal events. The current data is also informative with regard to how the presented scenes are temporally coded. Events that are processed as occurring concurrently are typically represented by more concrete, low-level event codes (Wakslak et al., 2006). If an event is represented as being not of one's 'direct concurrent experience', then representational codes tend to be more general or abstract. Wakslak et al. (2006) refer to events which have occurred in the past or are due to occur in the future as 'psychologically distant' and argue these types of event tend to be represented by high-level construals: Essential, abstract and global features of presented scenes. The more distal an event is considered to be, also depends on how far away from the observer that event is considered to be occurring, or whether the event is happening to the observer or to someone else.

If participants in the current experiment are generating representations of events which contain low-level construals (local features, concrete event codes), then it is likely that these events are also coded as occurring concurrently. More broadly, if events are considered to be ‘happening in front of us’ we are able to derive concurrent information from it and tend to process it using a concrete processing orientation (Wakslak et al., 2006). In addition, objects within scenes are often subject to differing processing orientations depending on how the event is temporally coded. For example, objects belonging to situations that are considered to be future occurrences tend to be represented using broader, type rather than token categories, compared to situations that are considered more proximal (Liberman et al., 2002; Liberman et al., 2007).

Grounded cognition and simulation. The data presented in the current chapter hold implications for several theoretical frameworks, and some tenets of these frameworks can be used to clarify the account of the data presented in this chapter. Where the current data serve to supplement these works, and where these works can inform the current account of the data will be discussed in the sections that follow. Firstly, Barsalou’s (2003) theory of ‘situated simulation’ assumes that category information does not form the main component of visual scene representations, but is instead used as a form of ‘scaffold’ for the current representation to graft event codes onto. This category information is not coded as general ‘type’ information of a particular scene, object or entity, but is instead thought to consist of re-enactments or ‘simulations’ of previously experienced events. These simulations are coded as sensory-motor states (defined as patterns of neural activation) and form part of a single representational domain, underlying both sensory-motor and conceptual processing. Importantly, representations of our visual environment are seen as dependent on context and highly dynamical. Sensory-motor simulations are fitted to the requirements of the current action-based situation. In other words, depending on one’s current goals or situational context, these simulations can change dynamically, rather than the same representation being activated over and over for each corresponding event.

Situated simulation, with its roots in grounded cognition (see Barsalou, 1993; 2008) is an apt theoretical construct to apply to the current data because of its focus on the interplay between perception and action. Grounded cognition does not view action and perception as being handled by separate, modular systems, but are instead viewed as residing within a single cognitive system which serves both. This common representational system makes use of combinations of partial re-enactments of sensory-motor states, visual percepts and conceptual processes.

As the current data presented in this chapter suggest, it is not a general representation of an event which is abstracted from the current scenes and subsequently used to facilitate recognition of the target. Rather, from a situated perspective, a sensory-motor state of the depicted event already held in the observer’s memory (as a result of previous experiential contact with the concurrent objects and derived knowledge as to the objects’ respective affordances) is activated. This is then used in conjunction with a representation of the concurrent scene, with this concurrent representation including low-level, global, and token-specific information.

Each observed event is not processed independently of available knowledge structure; additional knowledge and goal-based directives are brought to bear on the comprehension of the concurrent event. This notion is associated with ecological approaches to vision and the derivation of affordances from available objects and entities in our visual environment.

Ecological vision and affordances. Ecological vision provides a framework for the theory of affordances and the cognitive underpinnings of visual scene perception and object recognition. Ecological vision, as a precursor to more recent accounts of perception and action, forgoes the notion of representational cognition in favour of a more 'direct vision' approach (see Gibson, 1977). Put simply, this approach assumes one's perception of the world is derived directly from the senses, rather than from any internal representation. Nonetheless, more recent efforts have been made to tie some of the proposals of embodied cognition (a strongly representational field) with that of a central construct of the ecological approach – the affordance (e.g. Hirose, 2002).

The current data supplements the notion of the affordance as a theoretical construct: The affordances of instruments and objects in the depicted events are, as before, derived without the need of an agent to bring about the change in state of the target object. But, as mentioned previously, the affordances derived from stimuli in the current paradigm are likely specific to the objects presented. The affordances are not merely coded as whatever a particular category of object allows – knives usually afford bursting, but not in a specific case if the blade happened to be blunt. It is these specific details about the objects in question which serve to influence the nature of affordances derived from the event sequence.

Although the ecological framework upon which the notion of affordances rests represented an attempt to challenge representational theories (which had come to dominate the field at the time this approach was proposed), the current work is presented as cognitively representational in its descriptions and accounts of the data. Although the affordance can inform the account of the data presented here, observers' comprehension of visual scenes and recognition of objects in the current experiments is taken to require the construction and utilisation of appropriate event representations.

Theory of event coding. Finally, several ideas advocated by proponents of the theory of event coding (TEC) (Hommel et al., 2001) augment the findings presented in this chapter. This theory provides an account of the mental representation and coding of events in our visual environment, with a focus on perception and action planning.

TEC, in a core principle shared by grounded cognition, regards both perception and action to be managed by a common representational domain. It also makes use of the Gibsonian construct of affordances. However, this framework for understanding our computation of the world around us also utilises the concept of the 'feature code'. Feature codes refer to features of an event that can be local[5], as in the colour or shape of constituent objects within a

---

[5] Or 'proximal' to use Hommel et al.'s (2001) terminology, although this is not to be confused with the notion of proximal event codes discussed earlier in the chapter.

scene, or more complex, such as time and change, or the ‘sit-on-ableness’ of a surface (Hommel et al., 2001).

The notion of feature codes clarifies our account of the processes operating within the current paradigm as it offers an explanation as to the nature of information sharing within a single representational domain, capable of managing sensory-motor, perceptual, conceptual and action codes derived from perceived events. According to TEC, one event code can consist of a multitude of different feature codes, each representing a different facet of the perceived event. This notion of event codes being broken into many feature codes serves to provide a functionalised picture of the representational organisation of event codes.

Breaking down event codes into feature codes allows for suggestions as to how information is integrated and combined within a single representational store. For example, the notion of ‘feature binding’ has been proposed for explaining the synchronisation of separate, but supportive, feature codes (e.g. Abeles 1991; Singer 1994; Treisman 1996). Feature binding is a mechanism that integrates feature codes based on the same event. Separate features of an object (e.g. red, round) are selected and integrated to form ‘correct’ combinations of features from objects within a scene (Treisman, 1996).

Crucially however, any use of feature codes, feature binding or representational domains is ultimately required for action planning. In the paradigms presented and discussed within this thesis, facilitation of response times to a target can be thought of as one by-product of the ‘readying’ of certain representations for action planning and, of course, the resulting action. What links event comprehension and the retention of object-specific details is the need of feature codes for action planning. It is because neural architecture involved with action planning is activated upon seeing these ‘events’ that object-specific information is coded within the event representations in the current paradigm. In cases where no event structure can be abstracted from a visual scene (such as in change blindness studies), representations required for action planning are not being recruited. Action planning, although tangential to the current discussion, is the ‘next stage’ that is not accounted for by the current data.

Further research. Possibilities for future research focussing on the notion of object-specific representations reside in investigations of object affordances and graspability. As mentioned previously, the graspability of an object, even when no semantic information is associated with it, influences the types of event representation that can be generated from it.

If, as has been proposed in the account of the current data, event representations contain detailed causal information specific to the objects seen, then the paradigm can be constrained further to take account of the angle of some of the object surfaces. For example, previous discussion of graspability implies that if, in the current paradigm, the handle of the knife in the trial for the ‘burst’ concept was angled so as to be ungraspable by a right handed viewer (this was not methodologically controlled in the current experiment), priming effects would be eradicated as the affordances of the knife have been altered. Again, in this case knives still afford bursting, but this particular object’s arrangement does not allow the event to be so readily carried out.

A similar investigation of object angle was conducted by Yoon and Humphreys (2010). These authors found that right-handed participants were significantly faster to make a decision as to whether two objects commonly go together (such as a pan and a spatula), when these objects were held in appropriate hands by a protagonist (e.g. the spatula in the right hand and the pan in the left), than when they were not.

The proposed paradigm differs from this design, however, in that it moves away from functional groupings, using items which are not commonly found together and so there is less available semantic information for observers of the event to draw on (it is crucial to the current experiments that is it only with a particular combination of unrelated objects that priming arises). The proposed paradigm, therefore, is constrained further by including objects that are not semantic associates and also by excluding a depicted agent suitable for carrying out these events. The proposed paradigm would therefore seek to establish these types of effects operating in a highly constrained task environment.

This chapter has presented an experiment demonstrating priming effects with a distinction between same-token and different-token targets. In applying the notion of simulation to the current findings, a proposal for the types of mechanisms and knowledge structures involved in processing the stimuli in the current paradigm has been proposed. Sensory-motor states held in the observer's memory are used in conjunction with token-specific information derived from the depicted event. The current data also clarify the notion of the affordance, supporting subsequent representational theories that suggest a given object's affordances are not relegated to its category, but are malleable depending on the event context. The current chapter has sought to demonstrate, and then define, representational processes operating on available affordances in a paradigm which manipulates type-to-token changes between objects. Finally, the construct of feature codes applied within TEC elucidates our understanding of the current data by regarding the event code as broken up into several feature codes, each containing information specific to the objects observed within a concurrent event.

Chapter 6

Introduction

This chapter is intended, firstly, to provide an outline of background literature most pertinent to the current work, and to reiterate the main aims of the thesis. An overview of current empirical findings is then to be provided, with discussion of the theoretical implications of these results. And lastly, the limitations of the current work and potential future research directions are to be discussed.

There were two primary aims of the work presented in this thesis. The first of these aims was to establish whether it was possible to observe combinatorial processes in the visual domain similar to those observed in studies of language comprehension. The second was to establish the types of paradigmatic constraints under which these effects are borne out. These constraints include manipulating stimulus presentation times, the roles of concurrent objects, and the causal structure of visually presented events.

Language studies that established combinatorial semantic processing during sentence comprehension provided an impetus for the development of the current visual paradigms: Although the combining of semantic information from different words within sentences had been demonstrated in several studies (e.g. Tabossi, 1988; Potter & Falconer, 1979), a similar combining of semantic information from pictures of semantically unrelated objects had not been demonstrated.

Tabossi's (1988) study concerning the role of context in sentence processing determined that priming to certain words can be achieved with one particular combination of words within a sentence over another. For example, she found that participants were primed to the word ‘thorn’, upon seeing the sentence ‘the girl was pricked by a rose’. Here, the words ‘pricked’ and ‘rose’ are combined to form a representation of the event ‘pricking’. In this sentence, although the word ‘thorn’ is not used, it is primed based on the combined semantic knowledge engendered by the words ‘rose’ and ‘pricked’. Crucially, however, when participants were presented with sentences which do not contain the necessary combination of words, less priming was observed. For example, ‘thorn’ was not primed by the sentence ‘the girl smelled a rose’ because it did not contain the word ‘pricked’, and similarly this target was not primed from the sentence ‘the girl was pricked by a wasp’[6], as it did not contain the word ‘rose’.

In addition to semantic access, some paradigms have focussed on the abstraction of semantic information from collections of objects, encompassing a visual scene. Gist, typically considered to be the ‘general semantic interpretation’ of a visual scene (Henderson & Ferreira, 2004) can be abstracted in approximately 100 ms (Potter, 1976; Rensink, 2000). The rapidity with which gross semantic information can be gleaned form a visual scene is an indication that that this process operates early on in the processing of a visual scene, and that complete representations of visual stimuli do not need to be formed in order to ‘make sense’ of what we are seeing.

Priming effects from words or pictures and the abstraction of scene gist are now well established within the literature concerning sentence and scene comprehension. Important additions to these studies, however, have been those

---

[6] This original experiment was conducted in Italian and, as a result, some of the translated stimulus sentences sound slightly strange. In English, for example, one would not typically use the word ‘pricked’ here but ‘stung’ instead.

seeking to establish a link between perception and action. These experiments demonstrated priming, not to words or pictures, but to bodily actions, and were typically provided through a framework of embodied, or grounded cognition. Glenberg and Kaschak (2002) for example, developed several experiments that demonstrated an effect they termed the ‘Action-Sentence Compatibility Effect’ (ACE). The experiment in which the ACE was introduced had participants making sensibility judgements to sentences that described actions either towards the body (such as ‘open the drawer’), or away from the body (such as ‘close the drawer’). Participants were faster to make their judgements if, in responding, they needed to make a congruent arm movement. For example, responses to the sentence ‘open the drawer’ were faster if participants were required to make their response using an arm movement towards the body. These results are consistent with an embodied interpretation of visual processing in that the world around us is computed in terms of patterns of possible interactions (see Glenberg, 1997).

Recent methodology for exploring these notions of event representation, affordances, and the link between perception and action includes investigations of eye movements. Eye movements are used as, among other things, an indicator of semantic access and the activation of event knowledge. Typically, the ‘visual world paradigm’ tracks participants’ eye movements over a concurrent visual scene as they hear a sentence referring to elements of that scene. For example, Altman and Kamide (1999) presented participants with a scene in which a boy is sat beside a cake and a few other distractor items. These authors found that, as participants heard the sentence ‘the boy will eat the cake’, more saccades were made to the cake during the word ‘eat’, than any other item. However, when the word ‘eat’ was replaced with a different word, such as ‘move’, participants did not show any preference for an item within the scene during the acoustic lifetime of that word.

The authors concluded from this finding that, due to the cake being the only edible object in the scene, participants' eye movements reflected an anticipatory process based on the affordances of different objects within the scene. Specifically, the cake is the only item which affords eating, and so this item is anticipated on the basis that the boy will eat something in the scene. This eye movement experiment demonstrates the combining of affordances from two separate entities within a visual scene, based on the context offered by the auditory sentence. In this experiment, the affordances of the agent (the boy) are combined with the affordances of one of the most likely 'interactable' objects (the cake) based on our knowledge that an eating event is to occur.

Many eye movement studies provided an account of the interplay between vision, language and action, and it is maintained in this thesis that this interplay is crucial for action planning and conduction within our environment. The empirical studies reported in this thesis demonstrate combinatorial effects that are assumed to be an initial stage in the processing of a stimulus for future interaction.

Overview of empirical findings

The experiments presented in this thesis established combinatorial effects operating in the visual domain, demonstrated observers' sensitivity to causal structure within depicted events, and revealed the object-specific nature of the representations generated from objects presented within an event sequence.

Objects referred to in the context of the current set of experiments are entities upon which an instrument can instigate some form of change.

Experiments 1 and 2 explored the basic notion of combining semantic information from the affordances of different objects within a scene. Participants were presented with a sequence of pictures that in some cases suggested a particular event (such as a bird and a puddle for a ‘drinking’ event). Participants were significantly faster to make a lexical decision on an auditory word (e.g. ‘drink’) in this condition, compared to when the agent appeared without an appropriate theme (such as a bird with a book), when the theme appeared without an appropriate agent (such as a puddle and a book), or baseline unrelated objects were presented (such as a book and a toothbrush). When the exposure duration of first pictures in the sequence was shortened, this effect was eradicated.

These findings indicate sensitivity to the affordances of constituent objects within a scene. These results also suggest that it is only with a combination of items that is felicitous to a particular event (such as a bird and puddle for the event ‘drink’), that the event is primed. As these effects disappeared when shorter exposure durations of 100 ms were used, there may not have been enough time in these cases for event-based representations to be constructed from the available object affordances. In addition, these results suggest that, in order for priming to obtain from the visual domain to the linguistic domain, aspects of the visual representation (generated from the bird and puddle) must overlap with aspects from the language representation (generated from the target verb).

Experiments 3, 4 and 5 represented the first inclusions of causal information in the current paradigms, and served to constrain the paradigm further by using a purely visual task (an artifact judgement). In these experiments, participants were presented with ‘priming’ stimuli, consisting of an object and an instrument which is able to bring about some physical change in the object (e.g. a knife and a tyre), and ‘control’ stimuli, consisting of an object and an instrument which is unable to bring about a change in the object (e.g. a ruler and a tyre). Participants were required to make a judgement as to whether a subsequently presented target object (objects in their changed states such as a burst tyre) was man-made or not.

Participants were significantly faster to make artifact judgements in the prime, compared to the control condition. In order to establish the time-course of combinatorial effects and claim for automatic rather than strategic processes operating on these stimuli, the exposure duration of first presented pictures was reduced. Combinatorial effects in this case were severely diminished, with a significant difference between conditions observed only for by-items analyses.

These findings suggest that the event-based representations generated in these experiments contain information pertaining to the causal structure of the event. The causal structure of the depicted events can be thought of as the collected segments of an event which establish the role of respective items within a scene. For example, the knife and intact tyre suggest a future event of bursting, and the burst tyre represents the result of this event. These findings also suggest that combinatorial processes may not be solely automatic, but may require an element of strategic processing in order for them to operate.

Experiments 6, 7 and 8 sought to expand upon these findings by testing two basic assumptions of the previous experiments: That it is only with a felicitous combination of objects that priming to a related changed-state target occurs; and that participants are sensitive to the causal information contained within presented events. Participants in experiment 6 were presented with stimuli from four different conditions. In the first condition, items consisted of an object paired with an instrument capable of changing it (the basic priming condition used in experiments presented in previous chapters), and for the target, this same object in its end-state (as if altered by the instrument). The second condition consisted, again, of an object paired with an instrument which is able to instigate a change in the object, but an object in its *intact* state for the target (here, there is no change between the initial presented object and the target). The third and fourth conditions were identical to the first and second, except that the first presented pictures contained instruments on their own, without any objects. There were, however, no significant differences between any of the conditions in this experiment. A plausible explanation is that the previously reported effects were due at least in part to inhibition from the controls (see Neely 1991).

Experiment 7 reflected an effort to reduce the number of conditions due to the relatively small stimulus set used for experiments discussed so far. Experiment 8 explored the notion of inhibition from the control with a single first picture object. There were no significant results from these two experiments.

Finally, experiment 9 tested the assumption that the representations activated in the current paradigms were object-specific. Participants were presented with primes and controls (an object and instrument felicitous to an event, and an instrument and object infelicitous to an event, respectively), and were required to make an artifact judgement on changed-state targets. The first two conditions essentially contained the prime and control condition from the original ‘knife and tyre’ experiment (experiment 3), with the same changed-state objects from the targets (e.g. a burst tyre). The third and fourth conditions of this experiment contained exactly the same, except the changed-state targets appeared to be slightly different objects to those seen initially in the first presented picture, representing a different instantiation of the same event. Although this experiment contained four conditions, the stimulus set was doubled to maintain power in the experimental design. Using the example of the knife and the tyre, participants were primed to burst tyres but only to same-token targets (e.g. the same tyre as was seen in the first picture, although in its burst state). When responding to different- token targets (e.g. a burst tyre, but a different tyre to the one presented in the first picture), participants were unable to utilise the representation activated upon seeing the first picture (as it contains a different token of the same type) and so responses to the target in this case were facilitated much less.

Implications for vision

These findings indicate something fundamental about the way in which we process the world around us. Perhaps most importantly, these data suggest that semantic information is readily combined from objects in our visual environment so as to further comprehend our surroundings. Not only can these pieces of information be combined from objects that are not semantically associated (such as a knife and a tyre), but they appear to be combined based on what these objects afford in terms of some kind of interaction with it, or interaction with another object or entity. This abstraction of relevant information is closely tied to the notion of affordances: What an object means to us is what we can do to interact with it or what it can allow in order to interact with something else (Glenberg, 1997).

Although our perceptual experience of the visual world is one of a consistent, stable environment, our eyes make roughly 3-5 saccades per second, with very little information retained between each fixation (Rensink, 2000; Milner & Goodale, 1995). In addition, compete mental representations of visual surroundings are never formed. Instead, information is used from the environment as and when it is needed, there is no internally represented mental 'picture' of our environment to draw upon.

This notion of abstracting information from our visual environment 'cheaply' (Clark, 1999) can be expanded by the findings presented in this thesis, in particular its focus on combining affordances from separate objects to form a composite representation based on the event that, together, these items afford. From an event-based perspective, the notion of cheaply using available visual information is informative when one assumes the information abstracted from concurrent visual scenes aids both our perception of it and also our interaction with it. To clarify, much of our interpretation of the visual environment resides within our interface with it. It informs both perception and action, and affordances reflect a link between the two. As interaction with one's environment forms such an important part of our comprehension of the objects and entities around us, it is unsurprising that efficient visual processing encompasses the ways in which objects can be utilised for actions or events.

The likelihood that the visual system attempts to use available information as efficiently as possible suggests that, in doing so, it is not laden with irrelevant information that is not required to aid our perception of, and our interaction with, the visual environment. And, as evidenced by the work presented here, an important part of this efficient use of visual and semantic information is the combining of affordances from objects in our environment, aiding both the perception of visual scenes and our real-time interaction with them.

Another implication of the current findings is that a process common to the linguistic domain (combining of semantic information from separate sources) appears to operate in a similar way in the visual domain. This commonality between the two domains may reflect a single mechanism responsible for combining semantic information across any modality. As we typically experience things in our surroundings from several modalities, our comprehension of the external environment needs to be flexible and dynamic, requiring attention to several different sources, regardless of modality. The notion of a 'combinatorial' mechanism common to both linguistic and visual domains corresponds to Chase and Clark's (1972) assertion that in order to comprehend language or visual scenes, a *single* store of semantic knowledge is drawn upon (but see Paivio, 1986). The semantic priming literature contains several behavioural effects indicating the likely use of a single semantic store, utilised for both linguistic and visual stimuli (e.g. Vanderwart, 1984; Bajo & Canas, 1989). With regard to the current set of experiments, a combinatorial effect is demonstrated which bears striking similarity to effects reported in language studies which seek to establish processes involved with the comprehension and processing of text.

Implications for event representation

The representations generated from object affordances in the present experiments are rich enough to contain event-based information. Along with recognising individual objects within a scene, additional processing time allows for the generation of representations containing event information pertinent to the specific affordances of concurrent objects and instruments. The event-based information coded within these representations can be thought of as the types of events a given object or entity can engage in. Sensitivity to functional groupings of objects has been demonstrated before (e.g. Green & Hummel, 2006), but these current experiments indicate that representations concerning the interaction between two items can be generated purely from the affordances of these objects which, crucially, are semantically unrelated. For example, birds and puddles or knives and tyres are not semantically related pairs of objects, but they are linked in terms of what their respective affordances allow (drinking and bursting, respectively).

Another implication of the observed difference between the token-specific (same target) and the token-type (slightly different target) conditions is that the representations constructed from object affordances are specific to the objects seen and do not reflect global features of these objects' categories. In terms of the representations constructed from visual scenes, it is generally assumed that they include very little specific perceptual details of concurrent objects and entities (Grimes, 1996). Even when the demands of the task emphasise interactions, the information retained from the scenes is very simple, with basic spatial and object-type information coded within the representation

(Gilchrist et al., 1996). These findings suggest that the representations generated from these scenes contain general, abstract representational codes, much like those described in theories using such constructs as schemas or scripts (e.g. Ferretti et al., 2001; Rumelhart, 1980; Shank & Abelson, 1977). Although these representations are often said to contain object-type, rather than token-type codes, the representations engendered in the current experiments appeared to code token information within event structure. Also, the modulation of facilitation of responses to targets was clearly demonstrated based on an account of violations of causal structure within the depicted events. In experiment 9 (chapter 5), for example, same-token conditions (e.g. a knife and intact tyre for the first picture and the same burst tyre for the target) present cases where the causal structure of the event is constant. In these conditions, targets represent 'end-states' of the concurrent event, and instruments are attributed with causing the change seen in the target object. Conversely, in different-token conditions (e.g. a knife and tyre for the first picture and a different burst tyre for the target), the casual structure of the event is violated, and any representations generated from the initial picture are unable to facilitate recognition of the target.

Implications for grounded cognition

The abstraction of relevant information from visual scenes carries implications for the interaction between perception and action. The findings reported in this thesis support the Glenbergian notion that what an object means to us is 'what we can do with it' in terms of some kind of interaction (Glenberg, 1997). The data presented in this thesis emphasise the affordance as an important intermediary between our percept of a visual object and our subsequent interaction with it. The findings in this thesis provide an experimental account of the way in which objects and potential interactions and causal relationships between them are processed within our visual environment. One detail which resides outside the scope of this thesis is the potential mechanism responsible for combining semantic information from one object to the next. The current studies provide an adequate demonstration of this process at work, but they do not uncover any means for its occurrence. This shall be discussed later as a limitation of the current work.

For now, the abstraction of affordances is presented in this thesis as a critical stage in the processing of visual stimuli for subsequent interaction. Importantly, event representations are generated in the current set of experiments without the need of an agent that would be likely to carry out the particular action demanded of the event. As agents represent a common entity in depictions of events in virtually any medium, the fact that they are not needed for the construction of events involving an object and an instrument suggests that observers of the current visual scenes regard the stimuli, at least in part, as a concurrent event. In addition, the affordances of these instruments and objects are combined without the use of any semantic information linking the two. The two items are linked only in terms of what their combined affordances allow them to do in terms of a common event.

The current data suggest that, during our processing of visual stimuli, the affordances from constituent objects can be abstracted, and then combined and used for processing subsequent stimuli. Although this results in facilitation of the recognition of target objects, priming is attributed to the preparation of actions necessary for interaction with concurrent objects. This assertion is consonant with a core principle of TEC, in that perceptual processes are associated with action planning.

Lastly, the demonstration of object-specific representation activation in these current paradigms supports the notion of dynamic event codes that change depending on current goals or context (see Barsalou, 2003). From a situated perspective, object category information is assumed to consist of re-enactments of previously experienced events in the form of sensory-motor states, rather than general ‘type’ information. These representations are seen as highly dynamical and are readily altered so as to fit concurrent goals or events. More specifically, rather than a single ‘template’ representation being generated for each corresponding event, patterns of neural activity change dynamically according to the nature of the concurrent goal, action or event.

Implications for scene gist and object recognition

There are two major implications of the current data for gist abstraction. The role of gist in the current set of experiments can be separated into two possible scenarios. In the first case, gist abstraction may not be operating on the current stimuli at all. The affordances of different items within the scenes are combined and very little else occurs. Alternatively, gist abstraction may occur *before* the affordances of items within a scene can be combined.

Although gist abstraction differs in representational content to representations of object affordances, it is likely given the rapidity and automaticity with which gist is abstracted from scenes that this process operates upon the current stimuli. However, it is unclear at what point the combining of object affordances occurs, given the time needed for gist abstraction to take place.

If the combining of information from one object to another occurs in a faster, or similar time window to gist abstraction then there might not be enough 'room' for gist abstraction to take place. However, the combining of affordances from separate objects appears to take longer than the abstraction of gist from a visual scene. Experiments 2 and 5 both utilised shortened stimulus exposure durations, and both required the combining of information from presented objects. Combinatorial effects in these cases were severely weakened and, as we know, gist can be abstracted within 100 ms (Potter, 1975; 1976). These results suggested that the combining of affordances from these items takes longer to complete than was given to participants. It is unlikely therefore, that combinatorial processes are operating *instead* of gist abstraction in this time window.

Gist abstraction seems to be an automatic process, a gleaning of the 'identity' of a given scene, and likely occurs whenever such scenes are viewed. In early studies of gist abstraction, pictures were presented to participants at rates roughly similar to eye fixations. Even at such short exposure durations, participants were able to detect the same pictures in a later memory test or pick them out of a picture sequence. According to Rensink (2000) gist abstraction occurs early on in the processing of a visual stimulus, before stable representations of the stimulus can be constructed. Basic semantics inherent in the to-be-processed stimulus, referred to by Rensink as 'volatile structures', are used to glean the overall meaning of a scene from the available information, and one's attention can then be directed to objects of interest within the visual field.

These findings suggest that gist abstraction is a common process operating early on in the processing chain, with little to no strategic control. The fact that this process occurs in less than 100 ms reflects the automatic nature of this process (Potter, 1976). Thus, with regard to the current experiments, it is likely that gist abstraction is an initial stage in the processing of visual scenes, with affordance combining occurring later in the processing chain.

Another implication of the current data concerns the relative contributions of facilitation and inhibition in semantic priming paradigms. In the current experiments, the combinatorial effects appear to be based on both facilitation from primes and inhibition from controls. For example, experiment 7 (chapter 4) had participants responding in an artifact judgement task to conditions in which each instrument on experimental trials was capable of causing a change in the subsequently presented target object. Null results in this experiment were attributed to the lack of a control condition in which an instrument was shown that could not alter the target in any way. This was based on the assumption that priming effects were due, at least in part, to a 'mismatch' of representations generated from first pictures to targets, resulting in slower reaction times to the target.

Although this account appeared inconsistent given the results of subsequent experiments, this notion of inhibition from controls was raised again in relation to results obtained from experiment 9 (chapter 5). Here, participants were presented with either an instrument and object in its intact state, and the same object in its end-state for the target, or an instrument and intact object, but a slightly different token of the same type of object for the target. As predicted, participants were significantly faster to respond in an artifact judgement to targets that were the same as previously seen objects, only in their end-states, compared to targets which shared some, but not all of the features of the prior presented intact object.

Hommel's (1998) work concerning episodic binding suggests that repeating objects in semantic priming paradigms, but with slightly different features, causes 'code confusion' between two generated representations. Applying the notion of code confusion to experiment 9, a mismatch between representations generated from the presentation of a knife and intact tyre for the first picture, and a slightly visually different burst tyre for the target, may have induced inhibition of response times in this experiment. Unfortunately, it has proven difficult to establish with any degree of certainty, the relative contributions of facilitation and inhibition on a given priming paradigm (McNamara, 2005). However, this issue is somewhat elucidated by work suggesting that facilitation increases at shorter SOAs, while inhibition decreases at shorter SOAs (See Neely, 1991).

Limitations of the current work

With regard to the current work, there is one major theoretical limitation and three methodological limitations. Firstly, although combinatorial processes operating on visual scenes used in the current paradigms have been demonstrated and discussed clearly, a potential mechanism responsible for this type of processing has not been clearly defined. Secondly, the use of non-masked semantic priming paradigms restricts the type of claim that can be made concerning the time window in which this process operates. Thirdly, although efforts have been made to develop a purely visual paradigm, participants could not have been prevented from linguistically labelling the stimuli themselves, thus recruiting linguistic information in order to complete the task. And lastly, as participants were encouraged to remember presented stimuli, they could not have been prevented from developing mnemonic strategies in order to complete the task. Each of these limitations will be discussed, with consideration of potential improvements that could be made to the current paradigms.

The data presented in this thesis suggest that as our eyes move around our visual environment, the affordances of various objects are not only abstracted, but are combined to form representations that guide any interaction with our environment. It is maintained that this process of combining separate sources of semantic information is central to the planning and execution of actions. Moreover, affordance abstraction is established as a process that mediates between the perceptual and action planning systems. However, although a combinatorial process has been demonstrated in the current work, its description relies heavily on theoretical constructs such as the affordance and representation.

Unfortunately, although efforts could be made to localise this process in the brain using imaging methodology, this would still not give any clearer indication of the *cognitive* architecture directly responsible for such combinatorial effects. The current data are not sufficiently informative to allow consideration of such architecture and, therefore, an account of combinatorial processes based on a representational approach to visual semantics is offered in the current work.

In terms of the methodological limitations of the current experiments, the choice of an unmasked paradigm is perhaps one of the most significant. Visual pattern masks are often used in semantic priming paradigms to limit the amount of time available for processing the stimuli (see Wiens, 2006). Visual masks usually consist of a checkerboard or a similar pattern, and are typically presented between the offset of the prime and the onset of the target.

Pattern masks were not used in the current experiments because of concerns regarding the weakening of effects, and the effectiveness of such a method for constraining processing. Enns and Di Lollo (2000) contend that one of the main uses of pattern masking is not to restrict the amount of time spent on the processing of a particular stimulus, but is instead used to control the level of difficulty in the experimental task. As the effects reported in the current work appear in many cases to be fairly fragile (see chapter 4), increasing the difficulty of the task may have diluted any potential effects.

Another methodological concern with the current experiments is that, although the combinatorial effects established in the current experiments are assumed to operate upon purely visual stimuli, participants' utilisation of linguistic information to aid memory cannot be avoided. In an effort to ensure participants did not simply ignore the first picture in the two picture sequence, they were told that they would be undertaking a short memory test once the main phase of experimentation was over. There were no paradigmatic constraints ensuring participants in the current experiments were not labelling stimuli in an effort to retain the information they were presented with. Some previous studies required participants to tie processing capacity to an arbitrary task before responding to a target, such as counting aloud backwards (e.g. McNulty et al., 1994; Kelsey et al., 1999). However, if such a strategy were employed in the case of the current experiments, it would simply be replacing the use of an undesirable system in a visual task (e.g. linguistic systems), for another (e.g. numeric systems). In addition, along with constraining the types of systems that can be engaged in a particular task, using this method of tying up processing capabilities, the potency of the original priming may be diluted.

Another method to constrain processing would be to shorten stimuli exposure durations, pulling processing away from strategic to automatic processes (Sachs et al., 2008). However, as evidenced by current experiments in which shorter stimulus exposure durations were used, these combinatorial effects are weakened under a more constrained time window. Moreover, as work on representational cognition progresses, the view of the brain as housing discrete, modular systems is shifting in favour of a more distributed, linked system, where the influence of various cognitive systems may be inextricably linked and almost impossible to tease apart. Consequently, although a visual effect has been demonstrated that appears to be analogous to the combining of semantic information from different sources of linguistic information, the methodology involved with its demonstration is unable to rule out the use of linguistic information to complete the demands of the task.

Another possibility during these experiments is that participants were developing mnemonic strategies whilst viewing the experimental stimuli. Participants may have been encouraged to utilise such strategies because in each of the experiments they were expected to remember the stimuli as well as they could, as there would be a short memory test when the main experimental phase was over (see appendix 3). The influence of the use of mnemonic strategies in studies of memory, semantic priming, and sentence comprehension are numerous (Richardson, 1995; Corbett, 1977; Higbee, 2004), and serve to illustrate an alternative account, based on demand characteristics, of the priming observed in the current experiments.

The development of mnemonic strategies is related to the integrative account of the current data, discussed previously. The integrative account is presented as an alternative to the combinatorial semantic process assumed to occur during the processing of the stimuli. This account suggests that facilitation to targets is not caused by priming from the combination of items in the first picture, but that the target's altered state (e.g. a burst tyre) is confirmed given the items that precede it (e.g. a knife and a tyre). Crucially, this integration of items only occurs after the target has been presented.

A mnemonic strategy for remembering the initial objects (for later utilisation on the processing of the target) might involve consciously tying a story to each pair of items seen on the screen. For example, as mnemonics often rely on easily attributable associations between features or items, when presented with a knife and a tyre participants could tie a personal experience, amusing story, or rhyme to the observed objects (Higbee, 2004). It is, however, unclear whether there would be enough time for participants to develop these associations in the course of an experimental trial, for each pair of items seen.

While this issue cannot be ruled out as a potential contributory factor in the current experiments, participants' accuracy was taken to be an indication of improvement throughout the experiment. If mnemonic strategies were being employed by participants during the viewing of these stimuli, one would expect an improvement from the first to the second half of the experiment. However, although accuracy data was reported for each of the experiments, there were no observed improvement effects. Additionally, mnemonic techniques would be more effective for prime trials as the associations between primes and targets are more transparent. However, there were no differences between conditions in terms of the number of correct responses on experimental trials in any of the current experiments. These points provide evidence that mnemonic strategies were not used as a basis for experiment completion. In addition, the large number of fillers to experimental items ensured that the nature of the task was kept as obscure (barring the requirements of the experiment) to participants as possible.

Further research

There are two main avenues of future research. The first concerns the aforementioned contributions of automatic and strategic processing in the present paradigms. Although weaker combinatorial effects have been demonstrated for causal paradigms with shorter exposure durations (the ‘knife and tyre’ experiments with first picture durations of 250 ms), due to the fragile nature of the observed effects, the exact time course of these processes is still unclear. Additionally, stimulus exposure durations have not been manipulated for experiments with expanded stimulus sets. As combinatorial effects were stronger for later experiments that had better controlled conditions and twice as many stimulus items, timing manipulations could be used within the same paradigm to address this time course issue.

The current experiments could also be supplemented by studies seeking to explore the role of the graspability of objects within the current paradigms.

The angle of graspability of objects in the current experiments was not systematically controlled, although this has been established as a factor influencing affordance abstraction and the priming of actions. Yoon and Humphreys (2010), for example, found that participants make significantly faster decisions as to whether two objects ‘go together’ or not, when these objects were held by protagonists than when they were not. The current paradigms do not need agents to carry out the events, but current experiments could be altered to take account of the angle of graspability, and ensure it is consistent over all presented stimuli. Alternatively, an experiment could be developed in which the graspability of objects is manipulated, with right-handed participants responding to some stimuli arranged so as to be interactable by right-handed observers, and some to be interactable by left-handed observers. These suggestions and the proposed experiment above are the two main ways in which the work presented in this thesis could be developed further.

This thesis has presented findings seeking to establish combinatorial effects within semantic priming paradigms. The current experiments demonstrate observers’ sensitivity to the affordances of objects in a visual scene, and the combining of this semantic information to form relevant representations. These representations contain information regarding the event that is allowed by the combined affordances of previously presented stimuli. In addition, these effects have also been demonstrated operating in purely visual paradigms, and evidence has been provided for the object-specific nature of these event representations.

Appendices

Appendix 1 – Experiments 1 and 2 items

Target word	Picture 1	Picture 2
‘Break’	Boy with ball (prime)	Vase
	Umbrella (no agent)	Vase
	Boy with ball (no theme)	Umbrella
	Umbrella (baseline)	Rose
‘Build’	Boy with spade (prime)	Sandcastle
	Padlock (no agent)	Sandcastle
	Boy with spade (no theme)	Padlock
	Padlock (baseline)	Wheat
‘Catch’	Dog (prime)	Stick
	Sword (no agent)	Stick
	Dog (no theme)	Sword
	Sword (baseline)	Dolphin
‘Chase’	Cat (prime)	Bird
	Skull (no agent)	Bird
	Cat (no theme)	Skull
	Skull (baseline)	Rake
‘Chew’	Cow (prime)	Grass
	Mirror (no agent)	Grass
	Cow (no theme)	Mirror
	Mirror (baseline)	Waterfall
‘Drink’	Bird (prime)	Puddle
	Book (no agent)	Puddle
	Bird (no theme)	Book
	Book (baseline)	Toothbrush
‘Drive’	Farmer (prime)	Tractor
	Coat hanger (no agent)	Tractor
	Farmer (no theme)	Coat hanger
	Coat hanger (baseline)	Spider
‘Eat’ 1	Frog (prime)	Dragonfly
	Corkscrew (no agent)	Dragonfly
	Frog (no theme)	Corkscrew
	Corkscrew (baseline)	Hydrant
‘Feed’ 1	Horse (prime)	Woman hay
	Clock (no agent)	Woman hay
	Horse (no theme)	Clock
	Clock (baseline)	Paper
‘Gnaw’	Dog (prime)	Bone
	Keys (no agent)	Bone
	Dog (no theme)	Keys
	Keys (baseline)	Beehive
‘Juggle’	Juggler (prime)	Balls
	Hour glass (no agent)	Balls
	Juggler (no theme)	Hour glass
	Hour glass (baseline)	Wrench

‘Jump’	Man on horse (prime)	Small wall
	Sofa (no agent)	Small wall
	Man on horse (no theme)	Sofa
	Sofa (baseline)	Panda
‘Lift’	Woman (prime)	Child reach
	Bridge (no agent)	Child reach
	Woman (no theme)	Bridge
	Bridge (baseline)	Feather
‘Load’	Man with boxes (prime)	Van
	Fan (no agent)	Van
	Man with boxes (no theme)	Fan
	Fan (baseline)	Snake
‘Munch’	Boy (prime)	Crisps
	Rope (no agent)	Crisps
	Boy (no theme)	Rope
	Rope (baseline)	Pond
‘Peck’	Pigeon (prime)	Grain
	Pan (no agent)	Grain
	Pigeon (no theme)	Pan
	Pan (baseline)	Zebra
‘Play’	Cat (prime)	Ball of wool
	Toilet (no agent)	Ball of wool
	Cat (no theme)	Toilet
	Toilet (baseline)	Sunflower
‘Pull’	Child on sledge (prime)	Mother
	Torch (no agent)	Mother
	Child on sledge (no theme)	Torch
	Torch (baseline)	Raindrops
‘Push’	Woman (prime)	Child in pram
	Microphone (no agent)	Child in pram
	Woman (no theme)	Microphone
	Microphone (baseline)	Vegetables
‘Repair’	Mechanic (prime)	Car
	Tape measure (no agent)	Car
	Mechanic (no theme)	Tape measure
	Tape measure (baseline)	Maple leaf
‘Sell’	Seller (prime)	Fish
	Tree (no agent)	Fish
	Seller (no theme)	Tree
	Tree (baseline)	Ice
‘Serve’	Waiter (prime)	Buffet
	Pencil (no agent)	Buffet
	Waiter (no theme)	Pencil
	Pencil (baseline)	Balloon
‘Skip’	Girl (prime)	Skipping rope
	Basket (no agent)	Skipping rope
	Girl (no theme)	Basket
	Basket (baseline)	Lizard
‘Smell’	Girl (prime)	Flowers
	Boat (no agent)	Flowers

	Girl (no theme)	Boat
	Boat (baseline)	Screw
‘Spin’	Spider (prime)	Web
	Plunger (no agent)	Web
	Spider (no theme)	Plunger
	Plunger (baseline)	Molehill
‘Stroke’	Girl (prime)	Giraffe
	Pegs (no agent)	Giraffe
	Girl (no theme)	Pegs
	Pegs (baseline)	Peanut
‘Trap’	Spider (prime)	Fly
	Scissors (no agent)	Fly
	Spider (no theme)	Scissors
	Scissors (baseline)	Lilly pads
‘Wear’	Woman (prime)	Long coat
	Lamp (no agent)	Long coat
	Woman (no theme)	Lamp
	Lamp (baseline)	Pinecone

Appendix 2 – Experiment 3 bird-puddle and knife-tyre items

Bird-puddle items

Event	Picture 1	Picture 2
‘Blow’	Musician & trumpet (prime)	Whistle
	Zebra & trumpet (control)	
‘Bounce’	Man & basketball	Trampoline
	Giraffe & basketball	
‘Break’	Boy with a bat & window	Vase
	Woman with racket & window	
‘Build’	Boy & building blocks	Digger
	Stork & building blocks	
‘Carry’	Undertaker & coffin	Rucksack
	Ballerina & coffin	
‘Catch’	Dog & stick	Bb glove
	Rat & stick	
‘Chew’	Cow & grass	Sweets
	Gymnast & grass	
‘Cut’	Woman with knife & cake	Scissors
	Man with bottle & cake	
‘Drink’	Bird & puddle	Glass
	Boy & puddle	
‘Eat’	Frog & dragonfly	Plate
	Boy & dragonfly	
‘Hang’	Gardener & hanging basket	Noose
	Rhinoceros & hanging basket	
‘Lick’	Girl & ice cream	Stamp
	Peacock & ice cream	
‘Lift’	Mother & baby	Crane

	Toddler & baby	
‘Play’	Cat & ball of wool	Football
	Frog & ball of wool	
‘Push’	Mother & pram	Wheelbarrow
	Pig & pram	
‘Shake’	Mexican & maracas	Salt pot
	Soldier & maracas	
‘Spin’	DJ & turntable	Spinning top
	Baby & turntable	
‘Trap’	Spider & fly	Spring trap
	Dog & fly	

Knife-tyre items

Event	Picture 1	Picture 2
‘Bowl’	Bowling ball & pins	Scattered pins
	Wreath & pins	
‘Burst’	Knife & tyre	Burst tyre
	Ruler & tyre	
‘Crack’	Hammer & brick	Cracked brick
	Syringe & brick	
‘Cut’	Scissors & paper	Cut paper
	Glasses and paper	
‘Drink’	Straw & glass of juice	Empty glass
	Cigarette & glass of juice	
‘Inflate’	Pump & deflated balloon	Infl. balloon
	Bin & deflated balloon	
‘Iron’	Iron & crumpled shirt	Straight shirt
	Basket of flowers & crumpled shirt	
‘Mend’	Tape & broken broom	Mend. broom
	Tennis ball & broken broom	
‘Open’	Can opener & closed can	Open can
	Remote control & closed can	
‘Paint’	Paintbrush & fence	Painted fence
	Fly swat & fence	
‘Pour’	Bottle of wine & glass	Glass of wine
	Pepper mill & glass	
‘Screw’	Screw driver & panel without screw	Screw panel
	Dart & panel without screw	
‘Shoot’	Gun & target	Target holes
	Hairdryer & target	
‘Slice’	Bread knife & bread	Sliced bread
	Rolling pin & bread	
‘Smash’	Baseball bat & window	Smashed win.
	Rolled carpet & window	
‘Spray’	Aerosol can & wall	Sprayed wall
	Hourglass & wall	
‘Unlock’	Key & closed lock	Open lock
	Magnifying glass & closed lock	

‘Write’	Pen & blank sheet of paper Ladle & blank sheet of paper	Filled paper

Appendix 3 – Instructions for knife and tyre causal structure experiments

In this experiment, a cross will flash up on the screen (+) and you are required to move your eyes to it. Then, a picture will appear on the screen, before being replace with another cross. After the cross has disappeared a second picture will appear on the screen and you are required to make a decision as to whether you think the object on the screen is a man-made object or not.

If you think it IS man-made, then press the ‘yes’ key. If you think it is NOT man-made, please press the ‘no’ key.

Man-made objects will include things such as tables and chairs, musical instruments, household objects and so on. Objects that are not man-made might include such things as trees, rocks or outdoor scenes.

Please try to make this judgement AS QUICKLY AND ACCURATELY AS POSSIBLE.

After you press a button, the message ‘press key’ will appear. Just press the ‘yes’ button to go to the next trial – The first picture will appear and, just to remind you, when the second picture appears, please make a decision as to whether you think the picture shows something man-made or not. If you think it does press the ‘yes’ key and if you think it doesn’t press the ‘no’ key.

DON’T WORRY – There will be an opportunity to practice!
PRESS ANY KEY TO READ THE NEXT (AND FINAL) PAGE OF INSTRUCTIONS…

You can take a short break any time you see ‘press key’, but don’t forget to press ‘yes’ to advance to the next picture pair.

There will be a short practice session, and the experiment is quite short.

After the experiment is finished, there will be a second phase in which we shall test your memory for the pictures from the fist phase. Don’t worry, you will get further instructions about this after you have completed the experiment. For now, just try to make sure that you attend to the picture as you perform the task.

If you have any questions, please ask the experimenter, otherwise…
PRESS ANY KEY TO CONTINUE ONTO THE PRACTICE SESSION…

Appendix 4 – Debriefing information for knife and tyre casual experiments

THAT’S THE END OF THE EXPERIMENT

In this experiment, we wanted to look at whether we could encourage, or ‘prime’ your responses to the second picture in each trial, based on what you saw before it. On some trials, a picture would be presented to you which suggests a particular event (such as bursting, smashing, knocking and so on). We believe that these particular combinations of objects in these pictures will speed your reaction times to the picture that comes after it, if it is associated.

Actually, there is no memory test. You were told about this at the start because the experimenter was worried about participants simply ignoring the first picture.

Thank you for taking part.

Appendix 5 – List of items for experiment 4, including new and replaced items

D = different

Event	Picture 1	Picture 2
‘Assemble’	Glue & toy robot parts (prime a)	Assembled toy
	Ornament & toy parts (control a)	Assembled toy
	Glue & toy robot parts (prime b)	D assem. toy
	Ornament & toy parts (control b)	D assem. toy
‘Bind’	Elastic band & pencils	Bound pencils
	Cog & pencils	Bound pencils
	Elastic band & pencils	D bound pencils
	Cog & pencils	D bound pencils
‘Blow’	Dynamite & house	Blown-up house
	Cake & house	Blown-up house
	Dynamite & house	D blown-up house
	Cake & house	D blown-up house
‘Build’	Bucket with trowel & bricks	Built wall
	Bell & bricks	Built wall
	Bucket with trowel & bricks	D built wall
	Bell & bricks	D built wall
‘Burst’	Knife & tyre	Burst tyre
	Ruler & tyre	Burst tyre
	Knife & tyre	D burst tyre
	Ruler & tyre	D burst tyre
‘Carve’	Chisel & stone block	Carved block
	Lolly pop & stone block	Carved block

	Chisel & stone block	D carved block
	Lolly pop & stone block	D carved block
‘Chop’	Axe & chair	Chopped chair
	Bowl & chair	Chopped chair
	Axe & chair	D chopped chair
	Bowl & chair	D chopped chair
‘Clean’	Sponge & dirty window	Clean window
	Parcel & dirty window	Clean window
	Sponge & dirty window	D clean window
	Parcel & dirty window	D clean window
‘Crack’	Hammer & brick	Cracked brick
	Syringe & brick	Cracked brick
	Hammer & brick	D cracked brick
	Syringe & brick	D cracked brick
‘Cut’	Scissors & paper	Cut paper
	Glasses & paper	Cut paper
	Scissors & paper	D cut paper
	Glasses & paper	D cut paper
‘Demolish’	Wrecking ball & statue	Wrecked statue
	Boat & statue	Wrecked statue
	Wrecking ball & statue	D wrecked statue
	Boat & statue	D wrecked statue
‘Draw’	Pen & pad	Scribbled pad
	Spear & pad	Scribbled pad
	Pen & pad	D scrib. pad
	Spear & pad	D scrib. pad
‘Fill’	Hose pipe & watering can	Filled can
	Scarf & watering can	Filled can
	Hose pipe & watering can	D filled can
	Scarf & watering can	D filled can
‘Fuel’	Gas can & motorbike	Bike w fumes
	Perfume bottle & motorbike	Bike w fumes
	Gas can & motorbike	D bike fumes
	Perfume bottle & motorbike	D bike fumes
‘Hang’	Peg & shirt	Hung shirt
	Crisps & shirt	Hung shirt
	Peg & shirt	D hung shirt
	Crisps & shirt	D hung shirt
‘Install’	Battery & torch	Torch w light
	Packet of biscuits & torch	Torch w light

	Battery & torch	D torch light
	Packet of biscuits & torch	D torch light
‘Iron’	Iron & crumpled shirt	Straight shirt
	Basket & crumpled shirt	Straight shirt
	Iron & crumpled shirt	D str. shirt
	Basket & crumpled shirt	D str. shirt
‘Knock’	Sledge hammer & wall	Wall w hole
	Fishing net & wall	Wall w hole
	Sledge hammer & wall	D wall w hole
	Fishing net & wall	D wall w hole
‘Light’	Lighter & candle	Lit candle
	Sweet & candle	Lit candle
	Lighter & candle	D lit candle
	Sweet & candle	D lit candle
‘Mend’	Tape & broken broom	Mended broom
	Tennis ball & broken broom	Mended broom
	Tape & broken broom	D mended broom
	Tennis ball & broken broom	D mended broom
‘Open’	Secateurs & wire fence	Fence with hole
	Trowel & wire fence	Fence with hole
	Secateurs & wire fence	D fence with hole
	Trowel & wire fence	D fence with hole
‘Pack’	Clothes & open suitcase	Bulging suitcase
	Keyboard & open suitcase	Bulging suitcase
	Clothes & open suitcase	D bulging suitcase
	Keyboard & open suitcase	D bulging suitcase
‘Paint’	Wet brush & fence	Painted fence
	Fly swat & fence	Painted fence
	Wet brush & fence	D painted fence
	Fly swat & fence	D painted fence
‘Place’	Shoes & open shoebox	Closed shoebox
	Bone & shoebox	Closed shoebox
	Shoes & open shoebox	D closed shoebox
	Bone & shoebox	D closed shoebox
‘Polish’	Cloth & dirty ornament	Polished ornament
	Hat & dirty ornament	Polished ornament
	Cloth & dirty ornament	D pol. ornament
	Hat & dirty ornament	D pol. ornament
‘Pop’	Pin & balloon	Popped balloon
	Diary & balloon	Popped balloon

	Pin & balloon	D popped balloon
	Diary & balloon	D popped balloon
‘Scratch’	Knife & car	Scratched car
	Phone & car	Scratched car
	Knife & car	D scratched car
	Phone & car	D scratched car
‘Screw’	Screw driver and panel	Panel with screw
	Dart & panel	Panel with screw
	Screw driver and panel	D pan. with screw
	Dart & panel	D pan. with screw
‘Seal’	Small book & envelope	Sealed envelope
	Brick & envelope	Sealed envelope
	Small book & envelope	D sealed envelope
	Brick & envelope	D sealed envelope
‘Sharpen’	Stanley knife & pencil	Sharpened pencil
	Chocolate bar & pencil	Sharpened pencil
	Stanley knife & pencil	D sharp. pencil
	Chocolate bar & pencil	D sharp. Pencil
‘Shatter’	Cricket bat & mirror	Shattered mirror
	Coffee pot & mirror	Shattered mirror
	Cricket bat & mirror	D shattered mirror
	Coffee pot & mirror	D shattered mirror
‘Shoot’	Gun & bottle	Broken bottle
	Clip & bottle	Broken bottle
	Gun & bottle	D broken bottle
	Clip & bottle	D broken bottle
‘Shovel’	Trowel & litter	Pile of litter
	Radio & litter	Pile of litter
	Trowel & litter	D pile of litter
	Radio & litter	D pile of litter
‘Slash’	Knife & hanging bucket	Fallen bucket
	Brush & hanging bucket	Fallen bucket
	Knife & hanging bucket	D fallen bucket
	Brush & hanging bucket	D fallen bucket
‘Smash’	Baseball bat & window	Smashed window
	Rolled mat & window	Smashed window
	Baseball bat & window	D smash. window
	Rolled mat & window	D smash. window
‘Stack’	Fork-lift truck & boxes	Stacked boxes
	Bench & boxes	Stacked boxes

	Fork-lift truck & boxes	D stacked boxes
	Bench & boxes	D stacked boxes
‘Stitch’	Needle, thread & torn teddy	Stitched teddy
	Mask & teddy	Stitched teddy
	Needle, thread & torn teddy	D stitched teddy
	Mask & teddy	D stitched teddy
‘Sweep’	Broom & newspaper pages	Swept pages
	Door & newspaper pages	Swept pages
	Broom & newspaper pages	D swept pages
	Door & newspaper pages	D swept pages
‘Wrap’	Toy truck & wrapping paper	Wrapped present
	Paper bag & wrapping paper	Wrapped present
	Toy truck & wrapping paper	D wrapped present
	Paper bag & wrapping paper	D wrapped present

Appendix 6 – Experiment 3 norming questionnaire instructions

Bird and puddle norming instructions, version 1 (plausibility)

Please read the following instructions carefully before starting the experiment

---

Instructions

In this experiment, you will see either one or two objects and immediately below, a question asking how likely it is that something could happen. The following is an example:

How likely is shouting to occur?

1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○

If you think that shouting is highly likely given the mother and child, then you would select a high number on the scale. If you think it is unlikely to happen at all, then you would select a low number on the scale. A selection of ‘4’ means that it is no more or less likely for that thing to happen. A ‘7’ would mean that it would be almost certain to happen, and a ‘1’ that it would be almost certain it would not happen.

There is no right or wrong answer. We are simply interested in what you think about the likelihood of these events (e.g. climbing) given the objects shown. If the sentence had been ‘how likely is it that climbing would occur’, then it would make sense to give a lower ‘score’ to this case than to the shouting case, as shouting is more likely to occur here than climbing.
Questions in the experiment will look exactly as they do above, but with a large gap in between the picture and the selection buttons. There will also be two practice pictures before you start the experimental trials, and it shouldn't take any longer than 45 minutes to complete.

Your personal details

After the actual experiment ends, you will see a section asking for details about yourself. We would be grateful if you could give a valid email address so that we can contact you if we have any questions about your answers, and so that we can contact you if you have won the prize draw!

In the field marked "Region" we would like you to give us an indication of the region you grew up in, so that we have an idea of the type of English you speak (we would like this information in case there are differences between dialects!).

The personal data you give us is used only for scientific purposes. We will not give any of this information to anyone else (including other research groups), nor will we report any information in any way that can be identified with you.

And finally...

Taking part in this experiment is entirely voluntary! You should feel free to cancel the experiment at any point if you don't want to continue.

Once again, thanks for your interest in taking part. You can start the experiment proper by clicking on the start button below.

Start

Bird and puddle norming instructions, version 2 (likelihood, target)

Please read the following instructions carefully before starting the experiment

---

Instructions

In this experiment, you will be asked how likely it is that something will happen given the picture that appears immediately below the question. The following is an example:

How likely is it that drinking would occur?

1 ○ **2** ○ **3** ○ **4** ○ **5** ○ **6** ○ **7** ○

If you think that drinking is highly likely given this glass, then you would select a high number on the scale. If you think it is unlikely to happen at all, then you would select a low number on the scale. A selection of ‘4’ means that it is no more or less likely for that thing to happen. A ‘7’ would mean that it would be almost certain to happen, and a ‘1’ that it would be almost certain it would not happen.

For the above example, glasses are used for drinking and it would therefore be appropriate to select a high number (even a ‘7’). If, on the other hand, the sentence had read ‘how likely is it that biting would occur’, a sensible selection would have been ‘1’ as the act of biting given the glass is highly unlikely. There are no right or wrong answers. We are simply interested in what you think about the likelihood of these events (e.g. drinking) given the objects shown. Questions in the experiment will look exactly as they do above, but with a large gap in between the question and the picture. There will also be two practice pictures before you start the experimental trials, and it shouldn't take any longer than 45 minutes to complete.

Your personal details

After the actual experiment ends, you will see a section asking for details about yourself. We would be grateful if you could give a valid email address so that we can contact you if we have any questions about your answers, and so that we can contact you if you have won the prize draw!

In the field marked "Region" we would like you to give us an indication of the region you grew up in, so that we have an idea of the type of English you speak (we would like this information in case there are differences between dialects!).

The personal data you give us is used only for scientific purposes. We will not give any of this information to anyone else (including other research groups), nor will we report any information in any way that can be identified with you.

And finally...

Taking part in this experiment is entirely voluntary! You should feel free to cancel the experiment at any point if you don't want to continue.

Once again, thanks for your interest in taking part. You can start the experiment proper by clicking on the start button below.

Start

Bird and puddle norming instructions, version 3 (likelihood, agent)

Please read the following instructions carefully before starting the experiment

Instructions

In this experiment, you will be asked how likely it is that something will happen given the picture that appears immediately before the question. The following is an example:

`How likely is it that flying would occur?`

1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○

If you think that flying is highly likely given the bird, then you would select a high number on the scale. If you think it is unlikely to happen at all, then you would select a low number on the scale. A selection of ‘4’ means that it is no more or less likely for that thing to happen. A ‘7’ would mean that it would be almost certain to happen, and a ‘1’ that it would be almost certain it would not happen.

For the above example, birds tend to fly and it would therefore be appropriate to select a high number (even a ‘7’). If, on the other hand, the bird had been a penguin, then a ‘1’ would be more appropriate. Similarly, if the question had read ‘how likely is it that eating would occur’, you might want to give a lower number – although birds fly, eat, drink, and so on, eating isn’t the first thing that comes to mind when thinking of birds, whereas flying probably is (your average bird is probably more likely to fly off when you watch it than to eat). There are no right or wrong answers. We are simply interested in what you think about the likelihood of these events (e.g. flying) given the objects shown.
Questions in the experiment will look exactly as they do above, but with a large gap in between the picture and the selection buttons. There will also be two practice pictures before you start the experimental trials, and it shouldn't take any longer than 45 minutes to complete.

Your personal details

After the actual experiment ends, you will see a section asking for details about yourself. We would be grateful if you could give a valid email address so that we can contact you if we have any questions about your answers, and so that we can contact you if you have won the prize draw!

In the field marked "Region" we would like you to give us an indication of the region you grew up in, so that we have an idea of the type of English you speak (we would like this information in case there are differences between dialects!).

The personal data you give us is used only for scientific purposes. We will not give any of this information to anyone else (including other research groups), nor will we report any information in any way that can be identified with you.

And finally...

Taking part in this experiment is entirely voluntary! You should feel free to cancel the experiment at any point if you don't want to continue.

Once again, thanks for your interest in taking part. You can start the experiment proper by clicking on the start button below.

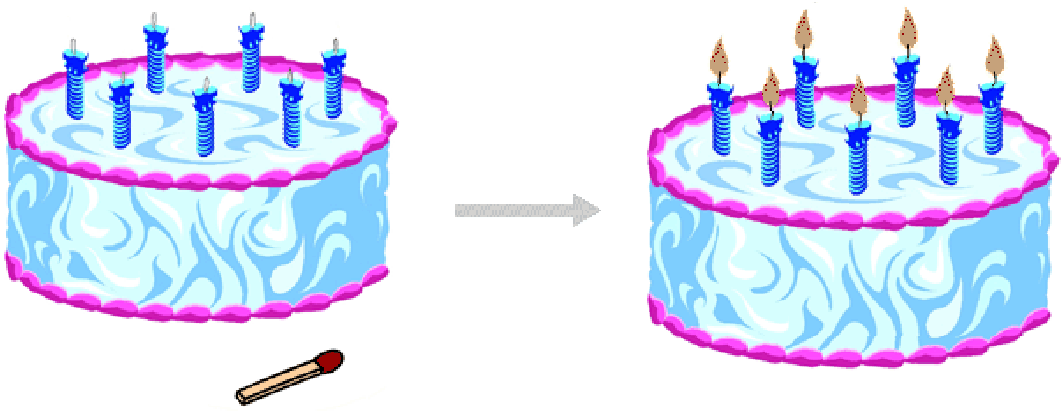
Start

Knife and tyre norming instructions, version 1 (plausibility)

Please read the following instructions carefully before starting the experiment

Instructions

Each question will show two pictures that suggest the ‘playing out’ of a particular event. For example, you might see the following:

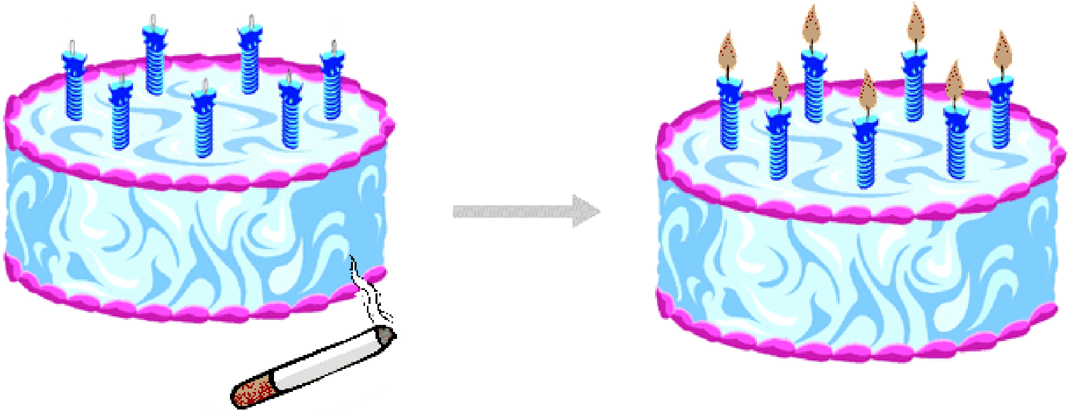


These pictures are intended to show an event in which a match is used to light the candles. Your task is to say how likely it is you would see this happening in everyday life. You can make your decision on a rating scale like this one:

1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○

If you think that lighting candles on a cake is highly likely to happen in everyday life, then you would select a ‘7’ on the scale. If you think it is a highly unlikely that this would happen, then you would select a ‘1’ on the scale. If you think this event is no more or less likely to happen in everyday life, then you would select a ‘4’.

There is no right or wrong answer. We are simply interested in what you think about the likelihood of these events happening. For example, you might have been presented with the following:



1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○

For this example, you may have wanted to choose a low score on the scale – the cigarette may have lighted the candles on the cake, but it is highly unlikely to see this in everyday life.

Your personal details

After the actual experiment ends, you will see a section asking for details about yourself. We would be grateful if you could give a valid email address so that we can contact you if we have any questions about your answers, and so that we can contact you if you have won the prize draw!

In the field marked "Region" we would like you to give us an indication of the region you grew up in, so that we have an idea of the type of English you speak (we would like this information in case there are differences between dialects!).

The personal data you give us is used only for scientific purposes. We will not give any of this information to anyone else (including other research groups), nor will we report any information in any way that can be identified with you.

And finally...

Taking part in this experiment is entirely voluntary! You should feel free to cancel the experiment at any point if you don't want to continue.

Once again, thanks for your interest in taking part. You can start the experiment proper by clicking on the start button below.

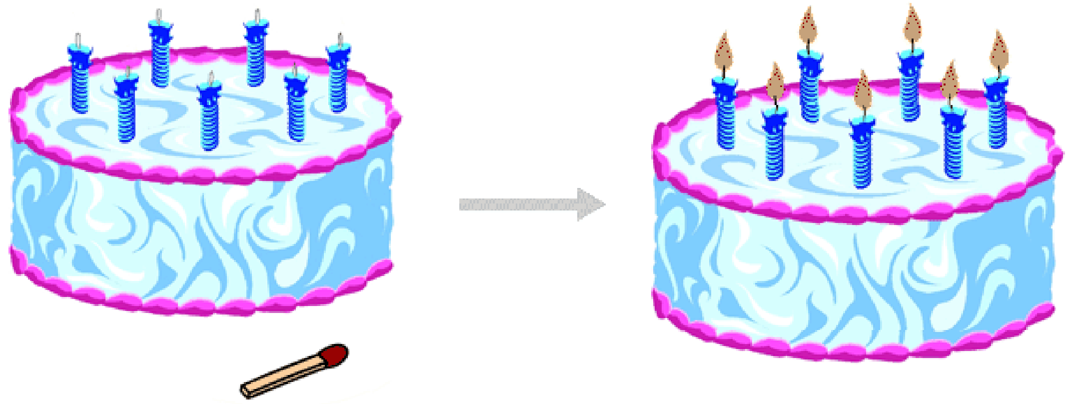
Start

Knife and tyre norming instructions, version 2 (likelihood, A and B)

Please read the following instructions carefully before starting the experiment

Instructions

Each question will show two pictures that suggest the ‘playing out’ of a particular event. For example, you might see the following:

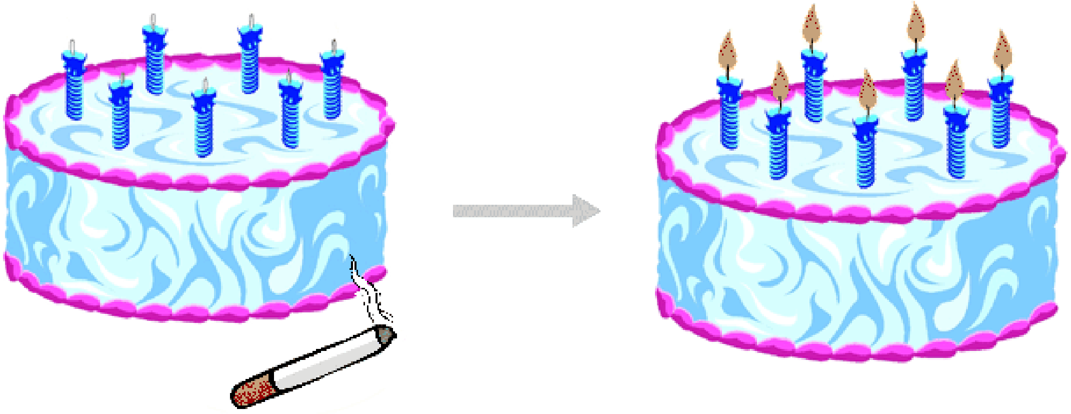


These pictures are intended to show an event in which a match is used to light the candles. Your task is to say how likely it is you would see this happening in everyday life. You can make your decision on a rating scale like this one:

1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○

If you think that lighting candles on a cake is highly likely to happen in everyday life, then you would select a ‘7’ on the scale. If you think it is a highly unlikely that this would happen, then you would select a ‘1’ on the scale. If you think this event is no more or less likely to happen in everyday life, then you would select a ‘4’.

There is no right or wrong answer. We are simply interested in what you think about the likelihood of these events happening. For example, you might have been presented with the following:



1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○

For this example, you may have wanted to choose a low score on the scale – the cigarette may have lighted the candles on the cake, but it is highly unlikely to see this in everyday life.

Your personal details

After the actual experiment ends, you will see a section asking for details about yourself. We would be grateful if you could give a valid email address so that we can contact you if we have any questions about your answers, and so that we can contact you if you have won the prize draw!

In the field marked "Region" we would like you to give us an indication of the region you grew up in, so that we have an idea of the type of English you speak (we would like this information in case there are differences between dialects!).

The personal data you give us is used only for scientific purposes. We will not give any of this information to anyone else (including other research groups), nor will we report any information in any way that can be identified with you.

And finally...

Taking part in this experiment is entirely voluntary! You should feel free to cancel the experiment at any point if you don't want to continue.

Once again, thanks for your interest in taking part. You can start the experiment proper by clicking on the start button below.

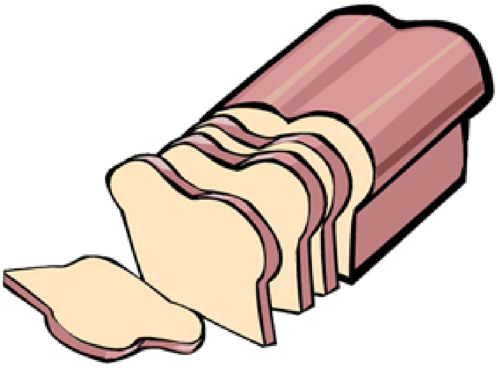
Start

Knife and tyre norming instructions, version 3 (commonality)

Please read the following instructions carefully before starting the experiment

Instructions

Each question will show a picture of an object. This picture is meant to represent an object in a particular state, rather than a 'kind' of that object. For example, an object may look squashed, broken, painted, etc. For instance, you might see loaf of bread that has been sliced:



Your task is to say how common you think a loaf of sliced bread is in everyday life. Note that we are asking you for a judgement about sliced bread, not bread in general, which in your experience could be more or less common than sliced bread.

You can make your choices on a rating scale like this one:

1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○

A choice of '7' means that you think sliced bread is 'very common', a choice of '1' means you think it is 'not very common at all', and a choice of '4' means you think sliced bread is no more or less common in everyday life than anything else.

Given an example to illustrate this, a book is very common, but a book with a torn page such as the one below is less common. If presented with this, you may want to choose a low score as this is not very common in everyday life. Remember, we are asking for a judgment on the specific state or condition of the object, not the class of objects.

There is no right or wrong answer. We are simply interested in how common you think these items are.

Your personal details

After the actual experiment ends, you will see a section asking for details about yourself. We would be grateful if you could give a valid email address so that we can contact you if we have any questions about your answers, and so that we can contact you if you have won the prize draw!

In the field marked "Region" we would like you to give us an indication of the region you grew up in, so that we have an idea of the type of English you speak (we would like this information in case there are differences between dialects!).

The personal data you give us is used only for scientific purposes. We will not give any of this information to anyone else (including other research groups), nor will we report any information in any way that can be identified with you.

And finally...

Taking part in this experiment is entirely voluntary! You should feel free to cancel the experiment at any point if you don't want to continue.

Once again, thanks for your interest in taking part. You can start the experiment proper by clicking on the start button below.

Start

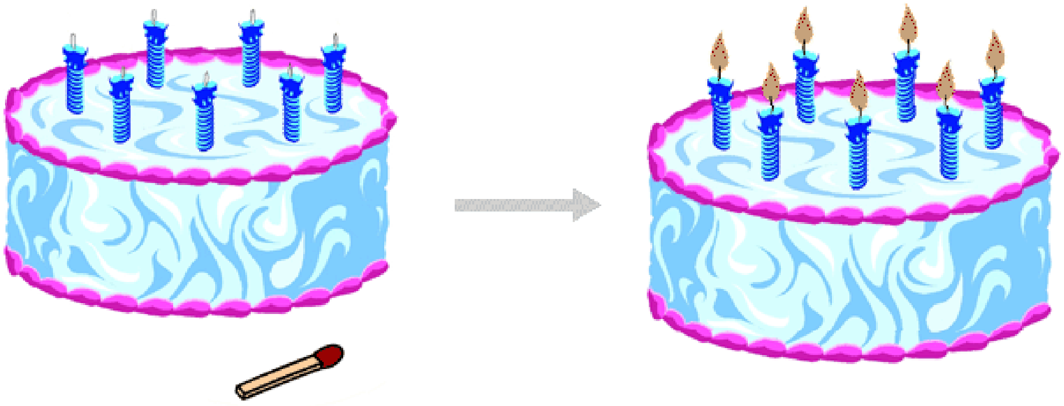
Appendix 7 – Experiment 4 norming questionnaire instructions

Experiment 4 norming instructions version 1 (plausibility)

Please read the following instructions carefully before starting the experiment

Instructions

Each question will show two pictures that suggest the ‘playing out’ of a particular event. For example, you might see the following:

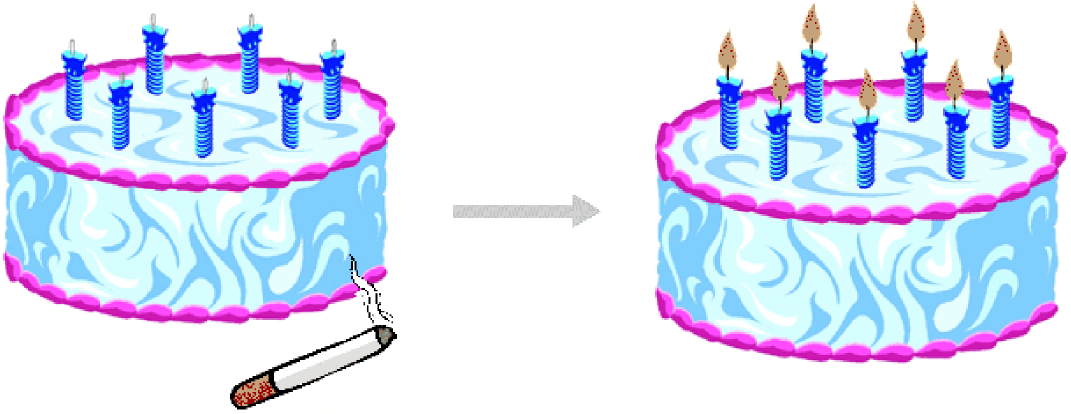


These pictures are intended to show an event in which a match is used to light the candles. Your task is to say how likely it is you would see this happening in everyday life. You can make your decision on a rating scale like this one:

1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○

If you think that lighting candles on a cake is highly likely to happen in everyday life, then you would select a '7' on the scale. If you think it is a highly unlikely that this would happen, then you would select a '1' on the scale. If you think this event is no more or less likely to happen in everyday life, then you would select a '4'.

There is no right or wrong answer. We are simply interested in what you think about the likelihood of these events happening. For example, you might have been presented with the following:



1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○

For this example, you may have wanted to choose a low score on the scale – the cigarette may have lighted the candles on the cake, but it is highly unlikely to see this in everyday life.

Your personal details

After the actual experiment ends, you will see a section asking for details about yourself. We would be grateful if you could give a valid email address so that we can contact you if we have any questions about your answers, and so that we can contact you if you have won the prize draw!

In the field marked "Region" we would like you to give us an indication of the region you grew up in, so that we have an idea of the type of English you speak (we would like this information in case there are differences between dialects!).

The personal data you give us is used only for scientific purposes. We will not give any of this information to anyone else (including other research groups), nor will we report any information in any way that can be identified with you.

And finally...

Taking part in this experiment is entirely voluntary! You should feel free to cancel the experiment at any point if you don't want to continue.

Once again, thanks for your interest in taking part. You can start the experiment proper by clicking on the start button below.

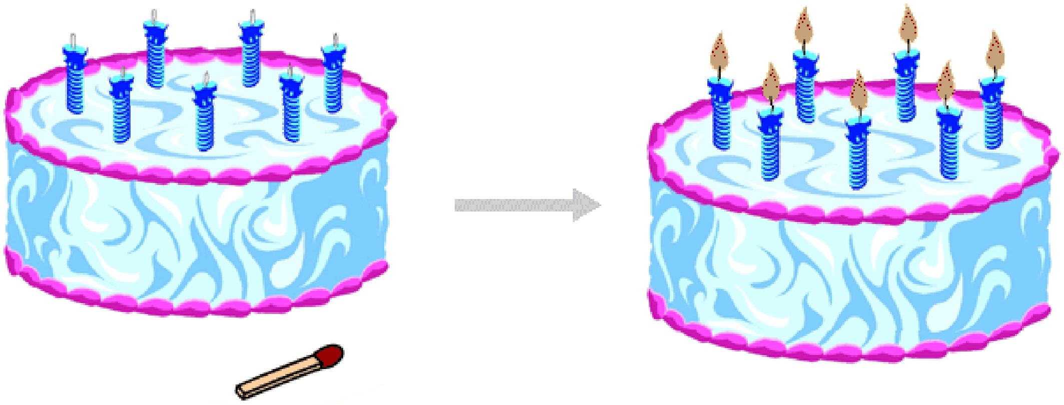
Start

Experiment 4 norming instructions, version 2 (likelihood)

Please read the following instructions carefully before starting the experiment

Instructions

Each question will show two pictures that suggest the ‘playing out’ of a particular event. For example, you might see the following:

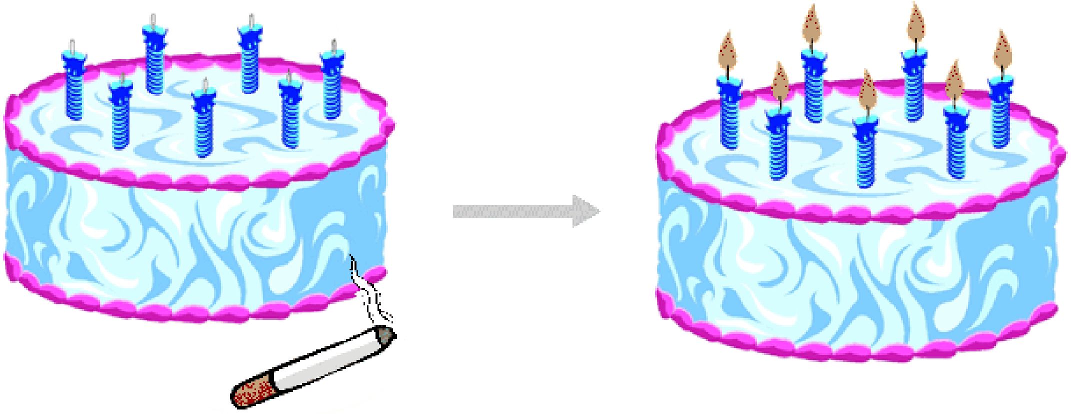


These pictures are intended to show an event in which a match is used to light the candles. Your task is to say how likely it is you would see this happening in everyday life. You can make your decision on a rating scale like this one:

1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○

If you think that lighting candles on a cake is highly likely to happen in everyday life, then you would select a ‘7’ on the scale. If you think it is a highly unlikely that this would happen, then you would select a ‘1’ on the scale. If you think this event is no more or less likely to happen in everyday life, then you would select a ‘4’.

There is no right or wrong answer. We are simply interested in what you think about the likelihood of these events happening. For example, you might have been presented with the following:



1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○

For this example, you may have wanted to choose a low score on the scale – the cigarette may have lighted the candles on the cake, but it is highly unlikely to see this in everyday life.

Your personal details

After the actual experiment ends, you will see a section asking for details about yourself. We would be grateful if you could give a valid email address so that we can contact you if we have any questions about your answers, and so that we can contact you if you have won the prize draw!

In the field marked "Region" we would like you to give us an indication of the region you grew up in, so that we have an idea of the type of English you speak (we would like this information in case there are differences between dialects!).

The personal data you give us is used only for scientific purposes. We will not give any of this information to anyone else (including other research groups), nor will we report any information in any way that can be identified with you.

And finally...

Taking part in this experiment is entirely voluntary! You should feel free to cancel the experiment at any point if you don't want to continue.

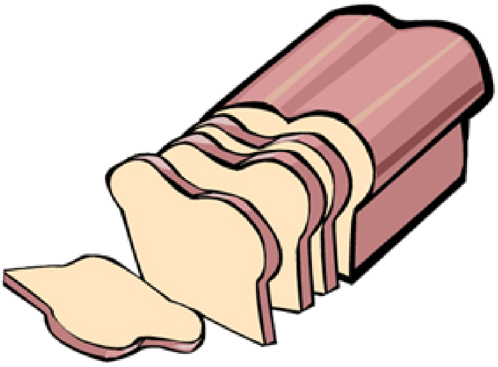
Once again, thanks for your interest in taking part. You can start the experiment proper by clicking on the start button below.

Experiment 4 norming instructions, version 3 (commonality)

Please read the following instructions carefully before starting the experiment

Instructions

Each question will show a picture of an object. This picture is meant to represent an object in a particular state, rather than a 'kind' of that object. For example, an object may look squashed, broken, painted, etc. For instance, you might see loaf of bread that has been sliced:



Your task is to say how common you think a loaf of sliced bread is in everyday life. Note that we are asking you for a judgement about sliced bread, not bread in general, which in your experience could be more or less common than sliced bread.

You can make your choices on a rating scale like this one:

1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○

A choice of ‘7’ means that you think sliced bread is ‘very common’, a choice of ‘1’ means you think it is ‘not very common at all’, and a choice of ‘4’ means you think sliced bread is no more or less common in everyday life than anything else.

Given an example to illustrate this, a book is very common, but a book with a torn page such as the one below is less common. If presented with this, you may want to choose a low score as this is not very common in everyday life. Remember, we are asking for a judgment on the specific state or condition of the object, not the class of objects.

There is no right or wrong answer. We are simply interested in how common you think these items are.

Your personal details

After the actual experiment ends, you will see a section asking for details about yourself. We would be grateful if you could give a valid email address so that we can contact you if we have any questions about your answers, and so that we can contact you if you have won the prize draw!

In the field marked "Region" we would like you to give us an indication of the region you grew up in, so that we have an idea of the type of English you speak (we would like this information in case there are differences between dialects!).

The personal data you give us is used only for scientific purposes. We will not give any of this information to anyone else (including other research groups), nor will we report any information in any way that can be identified with you.

And finally...

Taking part in this experiment is entirely voluntary! You should feel free to cancel the experiment at any point if you don't want to continue.

Once again, thanks for your interest in taking part. You can start the experiment proper by clicking on the start button below.

Start

References

Abeles, M. (1991). *Corticonics: Neural circuits of the cerebral cortex*. New York: Cambridge University Press.

Adamo, M., & Ferber, S. (2009). A picture says more than a thousand words: Behavioural and ERP evidence for attentional enhancements due to action affordances. *Neuropsychologia, 47*, 1600-1608.

Aloimonos, Y., & Rosenfeld, A. (1991). Computer vision. *Science, 253*, 1249-1254.

Altmann, G. T. M. Charles, V., & Gennari, S. P. (2005). Where language meets vision: event-structure in the visual world. *AMLap Conference, Ghent.*

Altmann, G. T. M., & Kamide, (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*, 247-264.

Anderson, M. L. (2003) Embodied cognition: A field guide. Artificial intelligence, 149, 91-130.

Andrews, S., Shelley, A. M., Fox, A. M., Catts, S. V., Ward, P. B., & McConaghy, N. (1993). Event-related potential indices of semantic processing in schizophrenia. *Biological Psychiatry, 34*, 443-458.

Bajo, M. T., & Canas, J. J. (1988). Phonetic and semantic activation during picture and word naming. *Acta Psychologica, 72*, 105-115.

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A. C. Collins, S. E. Gathercole, & M. A. Conway (Eds.), *Theories of memory* (pp. 29-101). London: Lawrence Erlbaum Associates.

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London: Biological Sciences, 358*, 1177-1187.

Barsalou, L. W. (2008). Grounding symbolic operations in the brain's modal systems. In G. R. Semin, & E. R. Smith (Eds.), *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches* (pp. 9-42). New York: Cambridge University Press.

Beeckmans, J. (2000). Generation of conceptual gists of abstract patterns. *Unpublished Article*, available at http://cogprints.org/1173/00/gist.htm.

Beer, A. L., & Diehl, V. A. (2001). The role of short-term memory in semantic priming. *Journal of General Psychology, 128*, 329-350.

Belke, E. (2008). Effects of working memory load on lexical-semantic encoding in language production. *Psychonomic Bulletin and Review, 15*, 357-363.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual Organisation* (pp. 213-253). Hillsdale, NJ: Erlbaum.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

Biederman, I., Blickle, T. W., Teitelbaum, R. C., Klatsky, G. J., & Mezzanotte, R. J. (1988). Object identification in nonscene displays. *Journal of Experimental Psychology: Human Learning, Memory, and Cognition, 14*, 456-467.

Block, N. (1983). Mental pictures and cognitive science. *Philosophical Review, 92*, 499-539.

Boyce, S., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 531-543.

Brown, M., & Besner, D. (2002). Semantic priming: On the role of awareness in visual word recognition in the absence of an expectancy. *Consciousness and Cognition, 11*, 402-422.

Campbell, F. W., & Wurtz, R. H. (1978). Saccadic omission: Why we do not see a grey-out during a saccadic eye movement. *Vision Research, 18*, 1297-1303.

Chase, W. G., & Clark, H. H. (1972). Mental operations in the comparison of sentences and pictures. In L. W. Gregg (Eds.), *Cognition in Learning and Memory*. New York: Wiley.

Chun, M. M. (1997). Types and tokens in visual processing: A double dissociation between the attentional blink and repetition blindness. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 738-755.

Clark, H. H. (1999). An embodied cognitive science? *Trends in Cognitive Sciences, 3*, 345-351.

Colcombe, S. J., & Wyer, R. S. (2002). The role of prototypes in the mental representation of temporally related events. *Cognitive Psychology, 44*, 67-103.

Corbett, A. T. (1977). Retreival dynamics for rote and visual image mnemonics. *Journal of verbal learning and verbal behaviou*r, 16, 233-246.

Craighero, L., Fadiga, L., Umiltà, C. A., & Rizzolatti, G. (1996). Evidence for visuomotor priming effect. *Neuroreport, 20*, 347-349.

Dell'Acqua, R., & Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition, 73*, 1-15.

den Heyer, K., Goring, A., & Dannebring, G. L. (1985). Semantic priming and word repetition: The two effects are additive. *Journal of Memory and Language, 24*, 699-716.

Doan, A., E. (2005). Type I and type II error. Encyclopaedia of social measurement, 883-888.

Dobel, C., Gumnior, H., Bölte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta Psychologica, 125*, 129-143.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4(9)*, 345-352.

Estes, Z. & Glucksberg, S. (1998). Contextual activation of features of combined concepts. Proceedings of the Cognitive Science Society, 333-338.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas and thematic role concepts. *Journal of Memory and Language, 44*, 516-547.

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General, 112*, 309-346.

Gajewski, D. A., & Henderson, J. M. (2005). The role of saccade targeting in the transsaccadic integration of object types and tokens. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 820-830.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Houghton-Mifflin.

Gilchrist, I. D., Humphreys, G. W., & Riddoch, M. J. (1996). Grouping and extinction: Evidence for low-level modulation of selection. *Cognitive Neuropsychology, 13*, 1223-1249.

Glaser, B. (1992). *Basics of grounded theory analysis*. Mill Valley, CA: Sociology Press.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20,* 1–55.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 558-565.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high dimensional and embodied theories of meaning. *Journal of Memory and Language, 43,* 379–401.

Grafton, S. T., Fadiga, L., Arbib, M. A., & Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *Neuroimage, 6*, 231-236.

Green, C. B., & Hummel, J. E. (2006). Familiar interacting object pairs are perceptually grouped. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1107-1119.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Eds.), *Perception* (Vancouver Studies in Cognitive Science, Vol. 5, pp. 89-109). New York: Oxford University Press.

Henderson, J. M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science, 8*, 51-55.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The interface of Language, Vision, and Action: Eye Movements and the Visual World*. New York: Hove, Psychology Press.

Higbee, K. L. (2004). Mnemonics, Psychology of. *International encyclopedia of the social and behavioural sciences*, 9915-9918.

Hirose, N. (2002). An ecological approach to embodiment and cognition. *Cognitive Systems Research, 3*, 289-299.

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition, 5*, 183-216.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*, 849-878.

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23*, 420-456.

Irwin, D. E. (1996). Integration information across saccadic eye movements. *Current Directions in Psychological Science, 5*, 94-100.

Irwin, D. E., Carlson-Radvansky, L. A., & Andrews, R. V. (1995). Information processing during saccadic eye movements. *Acta Psychologica, 90*, 261-273.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*, 175-219.

Kamide, Y., Atlmann, G. T. M., & Haywood, S. L. (2003). Prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133-156.

Kamp, H. & Partee, B. (1995). Prototype theory and compositionality. *Cognition*, 57, 129-191.

Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition, 27*, 117-143.

Kanwisher, N. G. (1991). Repetition blindness and illusory conjunctions: Errors in binding visual types with visual tokens. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 404-421.

Kelsey, R., M., Ornduff, S., R., Alpert, B., S. (2007). Reliability of cardiovascular reactivity to stress: Internal consistency. Psychophysiology, 44, 2, 216-225.

Koriat, A. (1981). Semantic facilitation in lexical decision as a function of prime-target association. *Memory & Cognition, 9*, 587-598.

Lavie, N., Lin, Z., Zokaei, N., & Thoma, V. (2009). The role of perceptual load in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1346-1358.

Liberman, N., Förster, J., & Higgins, E. (2007). Completed vs. interrupted priming: Reduced accessibility from post-fulfilment inhibition. *Journal of Experimental Social Psychology, 43*, 258-264.

Liberman, N., Sagristano, M., & Trope, Y. (2002). The effect of temporal distance on level of mental construal. *Journal of Experimental Social Psychology, 38*, 523-534.

Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., & Corbeille, J. L. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception & Performance, 33*, 1431-1450.

Mandler, J. M., & Parker, R. E. (1976). Memory for descriptive and spatial information in complex pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 38-48.

McKone, E. (1995). Short-term implicit memory for words and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1108-1126.

McNamara, T. P. (2005). *Semantic Priming: Perspectives from Memory and Word Recognition*. Taylor & Francis group: Hove, Psychology Press.

McRae, K., Hare, M., Ferretti, T. R., & Elman, J. L. (2001). Activating verbs typical agents, patients, instruments, and locations via event schemas. In *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 617-622). Mahwah, NJ: Erlbaum.

Michotte, A. (1963). *The perception of causality*. New York: Basic Books.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raso, V., & Rizzolatti, G. (1997). Objects representation in the ventral premotor cortex (Area F5) of the monkey. *Journal of Neurophysiology, 78*, 2226-2230.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory and Cognition, 4*, 648-654.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W.

Humphreys (Eds.), *Basic Processes in Reading: Visual Word Recognition*, pp. 264-336. Hillsdale, NJ: Erlbaum.

Neely, J. H., Keefe, D. E., & Ross, K. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 1003-1019.

Neisser, U., & Kerr, N. (1973) Spatial and mnemonic properties of visual images. *Cognitive psychology*, 5, 138-150.

Norris, J. (1984). The effects of frequency, repetition and stimulus quality in visual word recognition. *Quarterly Journal of Experimental Psychology, 36*, 507-518.

Mitchell, D. B., & Brown, A. S. (1988). Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 213-222.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*, 165-178.

Oliva, A. (2005). Gist of a Scene. In L. Itti (Eds.), *Neurobiology of Attention*, pp. 251-256. Elsevier, Academic Press.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences, 11*, 520-527.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory and Cognition, 3*, 519-526.

Peterson, R., & Simpson, G. (1989). Effects of backward priming on word recognition in single word and sentence contexts. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 1020-1032.

Posner, M. I., & Snyder, C. R. R. (1975b). Attention and cognitive control, In R. L. Solso (Eds.), *Information Processing and Cognition: The Loyola Symposion*. Hillsdale, NJ: Lawrence Erlbaum.

Potter, M. C. (1975). Meaning in visual search. *Science, 187*, 965–966.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 509-522.

Potter, M. C. (1999). Understanding sentences and scenes: the role of conceptual short-term memory. In V. Coltheart (Eds.), *Fleeting memories: Cognition of brief visual stimuli.* Cambridge, Massachusetts: MIT Press.

Potter, M. C., & Faulconer, B.A. (1979). Understanding noun phrases. *Journal of Verbal Learning and Verbal Behaviour, 18*, 509-521.

Potter, M.C., & Levy, E. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*, 10-15.

Rensink, R. A. (2000). Visual search for change: A probe into the nature of attentional processing. *Visual Cognition, 7*, 345–376.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1995). Image flicker is as good as saccades in making large scene changes invisible. *Perception, 24*, 26–28.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. P*sychological Science, 8*, 368–373.

Richardson, J. T. E. (1995). The efficacy of imagery mnemonics in memory remediation. *Neuropsychologia*, 33, 1345-1357.

Rossi, A. F., Desimone, R., & Ungerleider, L. G. (2001). Contextual modulation in primary visual cortex of macaques. *The Journal of Neuroscience, 21*, 1698-1709.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. E. Brewer (Eds.), *Theoretical Issues in Reading Comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sabb, F. W., Bilder, R. M., Chou, M., & Bookheimer, S. Y. (2007). Working memory effects on semantic processing: Priming differences in pars orbitalis. *Neuroimage, 37*, 311-322.

Sachs, O., Weis, S., Zallagui, N., Huber, W., Zvyagintsev, M., Mathiak, K., & Kircher, T. (2008). Automatic processing of semantic relations in fMRI: Neural activation during semantic priming of taxonomic and thematic categories. *Brain Research, 1218*, 194-205.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Earlbaum.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4*, 299-309.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5*, 195-200.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: some limitations of knowledge-based processing. *Cognitive Psychology, 14*, 489-537.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory and Cognition, 12*, 315-328.

Singer, W. (1994). Putative functions of temporal correlations in neocortical processing. In Koch, C. D. (Eds.), *Large-Scale Neuronal Theories of the Brain*. Massachusetts: MIT Press.

Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science, 7*, 301-305.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*, 261-267.

Simons, D. J., Franconeri, S. L., & Reimer, R. L. (2000). Change blindness in the absence of a visual disruption. *Perception, 29*, 1143-1154.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people in a real-world interaction. *Psychological Bulletin and Review, 5*, 644-649.

Symes, E., Ellis, R., & Tucker, M. (2007). Visual object affordances: Object Orientation. *Acta Psychologica, 124*, 238-255.

Tabossi, P. (1988). Effects of context on the immediate interpretation of unambiguous nouns. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 153-162.

Tabossi, P., Fanari, R., & Wolf, K. (2008) . Processing idiomatic expressions: Effects of semantic compositionality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34, 313-327.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology, 6*, 171-178.

Tsotsos, J. K. (1987). *Encyclopaedia of Artificial Intelligence*, (Ed.) S. Shapiro, John Wiley, New York, 389-409.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 830-846.

van Dam, W. O., & Hommel, B. (2010). How object-specific are object files? Evidence for integration by location. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1184-1192.

Vanderwart, M. (1984). Priming by pictures in lexical decision. *Journal of Verbal Learning and Verbal Behaviour, 23*, 67-83.

Versace, R., & Nevers, B. (2003). Word frequency effect on repetition priming as a function of prime duration and delay between prime and target. *British Journal of Psychology, 94*, 389-408.

Wakslak, C. J., Trope, Y., Liberman, N., & Alony, R. (2006). Seeing the forest when entry is unlikely: Probability and the mental representation of events. *Journal of Experimental Psychology: General*, 135, 641-653.

Warren, C., & Morton, J. (1982). The effects of priming on picture recognition. *British Journal of Psychology, 73*, 117-129.

White, P. A. (2006). How well is causal structure inferred from cooccurrence information? *European Journal of Cognitive Psychology, 18*, 454-480.

Wiens, S. (2006). Current concerns in visual masking. *Emotion*, *6(4)*, 675-680.

Wilding, J. (1986). Joint effects of semantic priming and repetition in a lexical decision task: Implications for a model of lexical access. *Quarterly Journal of Experimental Psychology, 38*, 213-228.

Yoon, E. Y., Humphreys, G. W., & Riddoch, M. J. (2010). The paired-object affordance effect. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 812-824.

Zacks, J. M., & Tversky, B. (2001). Even structure in perception and conception. *Psychological Bulletin, 127*, 3-21.