Behavioural Brain Research 236 (2013) 332–343

Contents lists available at SciVerse ScienceDirect

Behavioural Brain Research

journal homepage: www.elsevier.com/locate/bbr

Research report

Event-related potential correlates of emergent inference in human arbitrary relational learning

Ting Wang, Simon Dymond*

Department of Psychology, Swansea University, Singleton Park, Swansea SA2 8PP, United Kingdom

HIGHLIGHTS

- ► Examined functional–anatomical correlates of emergent relational inference.
- ► Relations trained between either words and pseudowords or arbitrary symbols.
- ► EEG was recorded during presentations of related and unrelated stimulus pairs.
- ► Faster, more accurate responses on symmetry and equivalence trials.
- ► ERPs were significant at mainly frontal–parietal and occipital sites.

ARTICLE INFO

Article history:
Received 2 July 2012
Received in revised form 17 August 2012
Accepted 21 August 2012
Available online 7 September 2012

Keywords:
Categories
Concepts
Relational inference
Stimulus equivalence
Relatedness
Directly trained
Event-related potentials

ABSTRACT

Two experiments investigated the functional–anatomical correlates of cognition supporting untrained, emergent relational inference in a stimulus equivalence task. In Experiment 1, after learning a series of conditional relations involving words and pseudowords, participants performed a relatedness task during which EEG was recorded. Behavioural performance was faster and more accurate on untrained, indirectly related symmetry (i.e., learn AB and infer BA) and equivalence trials (i.e., learn AB and AC and infer CB) than on unrelated trials, regardless of whether or not a formal test for stimulus equivalence relations had been conducted. Consistent with previous results, event related potentials (ERPs) evoked by trained and emergent trials at parietal and occipital sites differed only for those participants who had not received a prior equivalence test. Experiment 2 further replicated and extended these behavioural and ERP findings using arbitrary symbols as stimuli and demonstrated time and frequency differences for trained and untrained relatedness trials. Overall, the findings demonstrate convincingly the ERP correlates of intra-experimentally established stimulus equivalence relations consisting entirely of arbitrary symbols and offer support for a contemporary cognitive-behavioural model of symbolic categorisation and relational inference.

1. Introduction

Traditionally, psychologists and philosophers alike have been interested in the seemingly unique human propensity to classify, categorise and organise linguistic stimuli. Categorisation and concept formation abilities have long been considered to be defining features of symbolic behaviour that often cannot readily be traced to a history of direct learning. Early behavioural approaches, for instance, emphasised the role of principles of reinforcement, discrimination and generalisation in concept learning based on direct learning [1]. Recently, behavioural psychology has developed a fruitful and rigorous approach to the study of categorisation and symbolic behaviour, called derived relational responding, which is based on Sidman's [2] stimulus equivalence paradigm. Historically, the phenomenon of stimulus equivalence dates back to ancient Greece [3] and was studied by experimental psychologists, such as stimulus–response (S–R) theorists [4] for decades until the demise of S–R psychology [5]. However, it was not until the early 1970s that Sidman rediscovered the topic and set about devising a coherent set of experimental procedures and terminology with which to study it [6].

Research on stimulus equivalence and other forms of derived relational responding has generated considerable interest because it may provide a novel approach to the investigation of unlearned or emergent categorisation skills involving physically distinct, arbitrarily related stimuli. The basic finding shows that when verbally-able humans learn a series of interconnected conditional discriminations, the stimuli often become related to one another in ways not explicitly trained. For instance, if choosing Stimulus B in

* Corresponding author. Tel.: +44 1792 295602; fax: +44 1792 295679.
E-mail addresses: t.wang@swansea.ac.uk (T. Wang), s.o.dymond@swansea.ac.uk (S. Dymond).

0166-4328/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbr.2012.08.033

the presence of Stimulus A is taught (i.e., A–B), and choosing Stimulus C in the presence of Stimulus A (i.e., A–C) is also taught, it is likely that untrained relations will emerge between B and A, C and A (called *symmetry*), B and C, and C and B (called *combined symmetry and transitivity*, or *equivalence*), in the absence of any feedback. When these relations emerge, a stimulus equivalence relation is said to have formed among the relata [6,7]. These untrained, but nonetheless predictable, derived stimulus relations have been the focus of concerted research attention precisely because they are not readily explained by traditional behavioural principles of discrimination and stimulus generalisation. Neither B nor C, for instance, have a history of differential reinforcement with regard to each other (a defining feature of discrimination learning), therefore, neither should control selection of the other. Also, the outcomes cannot be accounted for on the basis of generalisation because the stimuli are all physically dissimilar or by simpler conditioning processes because that would involve an appeal to backward conditioning, which is weak at best [8].

The research interest generated by derived relations such as stimulus equivalence partly stems from the absence of an unequivocal demonstration of symmetry, transitivity, and equivalence in nonhumans [9]. There is minimal evidence of transitivity in rats [10], pigeons [11,12], chimpanzees [13], and monkeys [14]; of symmetry in rats [15], chimps [16], baboons and rhesus monkeys [17]; and widely contested, and as yet un-replicated, demonstrations of equivalence in chimpanzees [18] and sea lions [19]. The difficulties encountered obtaining positive test outcomes in nonhuman research on derived relations contrasts with the apparent ease with which humans, even young infants [20], have in passing tests for equivalence relations. This has lead many to propose that deriving stimulus relations may be a species-specific human ability underpinned by, or otherwise related to, language [9,21–23].

Indeed, contemporary behavioural theories of human language and cognition contend that derived relations such as stimulus equivalence may provide a new approach to studying symbolic behaviour and category formation because of the similarity between symmetry and equivalence test outcomes and the inherent bidirectional nature of word-referent relations in natural language [6,21,22]. According to these accounts, the relational nature of language, as studied with derived stimulus relations, provides the basis of modern accounts of language that overcome the limitations of previous approaches [21]. Yet, if research on derived relations is to inform contemporary behavioural accounts of language, then many of the effects seen in natural language research should be capable of being replicated with intra-experimentally acquired relations [21,24]. Support for this position comes from behavioural studies showing that reaction times to equivalently related word pairs are significantly faster than non-equivalently related word pairs [25]. Such findings show that equivalence relations may produce effects similar to semantic priming effects observed in the cognitive literature when real words are used [26].

Related research on the neural mechanisms underlying equivalence relations has recently begun to identify potential overlap with brain regions responsible for language and semantic processing. For instance, neuroimaging studies highlight a key role for frontal–subcortical and frontal–parietal networks in the emergence of derived relations [27–31]. One study [27] found activation in inferior frontal (dorsolateral) and inferior parietal regions for both trained (i.e., A–B and B–C) and derived (i.e., B–A, C–B, A–C, and C–A) relations. Schlund et al. [28,29] extended these findings by conducting event related analyses of BOLD signal change correlated with blocks of trained, derived relations and matching control relations. They found that accurate responding on trained and derived trials was correlated with bilateral activation in inferior parietal lobule, dorsolateral and ventrolateral inferior frontal regions, and in area of the thalamus and globus pallidus. Activation in the inferior parietal region during stimulus equivalence tasks is consistent with studies that have examined transitive inference and other forms of inference-based relational learning [32,33]. The recruitment of the hippocampus during transitivity and equivalence relations, and the demonstration of activation in the parahippocampus to cross-class (unrelated) control tasks further reveal hippocampal involvement in relational inference tasks such as tests for stimulus equivalence relations that do not rely on a serial order structure [29]. Recently, prefrontal, medial frontal, and intraparietal cortices were activated during tests for symmetry, transitivity and equivalence, with additional activation in the precuneus and posterior parietal cortex during transitivity and equivalence testing [30]. Taken together, neuroimaging studies of stimulus equivalence indicate a role for a distributed frontal–parietal, hippocampal system in integrating and processing non-adjacent stimuli in a manner resembling that reported by previous studies using similar tasks [28,33].

To date, only a handful of event-related potential (ERP) studies have used the stimulus equivalence paradigm to determine the functional–anatomical correlates of untrained, inferential categorisation. In the first such study, Barnes-Holmes et al. [34] sought to examine whether the N400 ERP [35], evoked by semantic incongruity, was modulated by related and unrelated non-words learned intra-experimentally. First participants were exposed to training with a series of conditional relations involving non-existing words ('pseudowords', referred to here for the purposes of clarity with alphanumerics: A1–B1–C1–D1, A2–B2–C2–D2; participants were never exposed to these labels) until a high mastery criterion was achieved (M trials completed = 398). Next, pairs of stimuli were presented while EEG was recorded in a two-word lexical decision task with combinations of class–class (i.e., directly trained [A1–B1], symmetry [B1–A1], transitivity [A1–C1], and equivalence [C1–A1]) and class–nonclass trials (i.e., A1-C2), along with trials involving novel nonsense words (class–nonsense, nonsense–class, and nonsense–nonsense). Stimuli consisted entirely of pronounceable, orthographically regular six-letter pseudowords [36]. Reaction time and error data showed participants responded significantly faster and made fewer errors with directly and indirectly related pairs than to unrelated pairs. Generally, ERPs revealed a greater negativity indicative of the N400-like effect during unrelated pairs (i.e., class–nonclass and novel trial types) than to directly trained or equivalent pairs. The directly trained, equivalent, and nonequivalent trials selected for analysis by Barnes-Holmes et al. [34] closely resemble the directly related, indirectly related, and unrelated word pairs often examined in ERP studies of semantic priming [26,37] and provide preliminary support for a derived relations model of semantic-like processing of pseudowords.

A further study [38] compared ERPs to pairs of identical stimuli (i.e., reflexivity trials) and related and unrelated equivalence stimuli presented in a two-choice matching to sample task. Occipital P2 and frontal N2 components were evoked during reflexivity trials (where the matched stimuli are identical), while a later component, a parietal P3, was observed during equivalence test trials. Subtracted ERP components (unrelated-reflexivity and unrelated-equivalence) revealed a significant dN400-like effect with similar scalp topography to that reported in previous studies on stimulus equivalence and relational matching [34,39,40], thus providing further evidence of semantic-like processing during tests for equivalence.

A final study on the ERP correlates of stimulus equivalence [41] compared related vs. unrelated stimulus pairs in small groups of participants who received EEG recording either before or after matching to sample tests for equivalence. Only those participants who received EEG recording after equivalence tests showed an N400-like effect, as revealed by difference waves (unrelated-related). These findings are limited, however, by the small sample