Paper:
Quigley, M. & Reed, P. (2017). Over-selective Responding in a Diagnostic Judgment Task. *Applied Cognitive Psychology, 31*(5), 558-564.
http://dx.doi.org/10.1002/acp.3341

Over-selective responding in a diagnostic judgment task

Martyn Quigley and Phil Reed

Swansea University, UK

Correspondence address: Phil Reed,

Department of Psychology,

Swansea University,

Singleton Park,

Swansea, SA2 8PP, U.K.

e-mail: p.reed@swansea.ac.uk

There was no external financial support for this study.

Short title: Over-selective responding.

Cite as: Quigley, M., & Reed, P. (2017). Overselective responding in a diagnostic judgment task. *Applied Cognitive* Psychology, **31**, 558-564. 10.1002/acp.3341.

Abstract

Medical diagnoses are often made based on the presence of multiple-symptoms. However, little is known about how the presence of multiple simultaneous-symptoms may influence bias in determining which symptoms are identified, in part due to a lack of an experimental analogue of this process. The current manuscript presents a laboratory analogue of this process, and explores whether over-selectivity influences the ability to identify symptoms indicative of particular illnesses. In two experiments, participants completed a diagnosis task which required them to rate the degree to which symptoms predicted illnesses, with predictor symptoms being presented either singly or in compound. In both experiments, over-selectivity was observed; one symptom of the compound received lower ratings, and lower ratings than the single predictor, whilst the other component received comparable ratings to the element. These findings are discussed in relation to associative accounts of over-selectivity and as a procedure to study biases in medical-decision making.

Key words: experimental analogue; over-selectivity; judgment; medical decision making, diagnosis.

'Over-selectivity' refers to the phenomenon whereby one stimulus from a set of equally valid predictors of an outcome comes to control behavior (e.g., Dube & McIlvane, 1999; Lovaas, Koegel, & Schreibman, 1979; Reed & Gibson, 2005). Over-selective responding has been noted to be a problem for a wide range of different populations, including those with an autism spectrum disorder (Allen & Fuqua, 1985; Huguenin, 1997; Leader, Loughnane, McMoreland, & Reed, 2009; Lovaas & Schreibman, 1971), an intellectual disability (Dube & McIlvane, 1999; Stromer, McIlvane, Dube, & Mackay, 1993), the elderly (Kelly, Leader, & Reed, 2016), and those with some forms of brain injury (Wayland & Taplin, 1985). However, it has also been shown in healthy typically-developing individuals who are operating under conditions of cognitive load; that is, the participants are engaged in a second task while they are also completing the 'over-selectivity' task (Dube, Balsamo, Fowler, Dickson, Lombard & Tomanari, 2006; Reed, 2006; Reed & Gibson, 2005; Reynolds & Reed, 2011).

Over-selective responding has been linked to disruptions in the ability to learn complex tasks requiring the conjoint processing of several pieces of information (Chiang & Carter, 2008; Koegel, Schreibman, Britten, & Latinen, 1979), and can be seen as an experimental analogue of decision making, when multiple pieces of information are presented and require processing (e.g., Keinan, 1987). Developing an understanding of the relationship between over-selectivity and decision-making under these conditions may have important implications for processes like medical diagnoses which are often made as a result of the presence of multiple, potentially competing symptoms These circumstances have been suggested to lead to some symptoms being missed by health professionals, potentially causing misdiagnosis, or at least problems in identifying an illness (e.g., Maserejian, Link, Lutfey, Marceau, & McKinlay, 2009; Resnick, Brandeis, Baumann, DuBeau, & Yalla, 1996). The medical diagnosis process has a face-similarity to the manner in which the presentation of multiple stimuli can induce over-selectivity in the laboratory. If a stronger link between

these two areas could be established, then this might allow insights about the conditions driving over-selective responding derived from the laboratory to be applied to such real-world decision making situations, to illuminate further our understanding of these important issues.

Often the conditions under which over-selectivity emerge are studied using a discrimination learning task. Participants are initially exposed to a simultaneous discrimination task between two compound stimuli, of the form AB+ CD- (e.g., Reed & Gibson, 2005; Reynolds & Reed, 2011). The participants are presented with two compounds, AB which is reinforced (+), and CD which is not reinforced (-). Once they have learned the discrimination to a criterion, they are presented with a number of test trials in extinction, in which one element from the previously reinforced compound and one element from the previously non-reinforced compound are presented together (e.g., AvC; AvD; BvC; BvD). Those showing over-selective responding (e.g., those with ASD, those with a cognitive load) will typically choose one previously reinforced element (e.g., A) to a greater extent than the other (e.g. B) on these trials, whereas controls will choose A and B equally often.

For example, Reynolds and Reed (2011) presented participants with two cards each containing two black-and-white hand drawn images, taken from the British Picture Vocabulary Scale. Prior to starting the experiment participants were asked to point to one of the compound sets (i.e. one of the cards containing two images) and informed that they would be provided with corrective verbal feedback after they had made their selection. If participants selected the reinforced compound set, AB+ (e.g. a bus and an elephant), they would be told “yes” by the experimenter. If, however, participants pointed to the non-reinforced compound set, CD- (e.g. a leaf and a hat) the experimenter would say “no”. Once participants had responded correctly to the reinforced compound (AB+) for ten consecutive trials they would proceed to a test stage. In this stage participants would be presented with one element (e.g. elephant) of the reinforced compound and one element of the non-

reinforced element (e.g. leaf), and this would occur for each of the stimuli presented in the training stage (e.g. AvC; AvD; BvC; BvD); importantly participants would receive no verbal feedback during this stage. It was found that participants selected one element of the AB compound at test (e.g. elephant) more frequently than the other element (e.g. bus). Thus demonstrating over-selective responding.

Although the range of populations that have been shown to display over-selective responding is great, however, the range of procedures that have been used to study it are limited, and are heavily based on conditioning studies (see Reynolds & Reed, 2011). It is, therefore, not clear whether these basic procedures could inform understanding of the processes occurring in more complex tasks, such as those discussed above in relation to medical decision-making (e.g., Maserejian et al., 2009; Resnick et al., 1996). Given the potential theoretical insights that such laboratory procedures might bring to these applied areas, the current study aimed, simply, to extend the investigation of over-selectivity to a paradigm that is more related to the study of medical decision making (in this case, diagnosis).

To this end, a variant of a causal judgment task (i.e. a diagnostic judgement task) was utilised to explore whether an over-selectivity effect would emerge. In a judgement task a number of stimuli serve as causes of a particular event or outcome, and participants are asked to make judgements about the likelihood of the event or outcome occurring as a result of the different stimuli. For instance, in a commonly used judgement task, Van Hamme and Wasserman, presented participants with a number of foods - some of which would cause an allergic reaction - and participants were asked to rate to what extent the food caused an allergic reaction. In the present diagnostic judgement task, however, symptoms predicted fictitious illnesses and participants were required to identify which symptoms indicated the presence of which illness, in order to provide an experimental analogue for medical decision making. Participants were presented with both an elemental stimulus (e.g. A = ‘Skin Rash’),

and a compound stimulus (e.g. BC = ‘Stomach Ache’, ‘Blurred Vision’), that comprised potential symptoms that predicted fictitious illnesses. The elemental stimulus always predicted one illness, whilst the compound stimulus always predicted an alternative illness. If over-selective responding were to emerge in this procedure, then one of the symptoms in the compound cue should be rated as predicting illness to a greater extent than the other element; i.e. one symptom would be seen as controlling the diagnosis at the expense of the other. Given the limited range of procedures under which the effect has currently been observed, the presence of the effect in the present study would provide further evidence of its generality and highlight its implications when applied to important real world issues such as medical decision making. In addition, the degree to which the elements from the compound would be rated as illness-predictive was also compared to the ratings given to an elemental predictor symptom. This comparison was made to see if the presence of several symptoms would reduce the degree to which all are learned about, as predicted by some key theories of learning (see Rescorla & Wagner, 1972), but not by all (see Mackintosh, 1976). Importantly, these theories of learning could provide an insight into the mechanisms contributing to over-selective responding.

Experiment 1

In Experiment 1, a diagnostic judgment task was employed that required participants to make a rating about how predictive various symptoms of two different fictitious illnesses were, in order to explore whether the over-selectivity effect would emerge in this context. Participants were presented with an elemental stimulus (A), and a compound stimulus (BC), that each predicted different fictitious illnesses. The elemental stimulus always predicted one illness, whilst the compound stimulus always predicted an alternative illness. A further three non-target stimuli (D, E, F) were also presented which predicted no illness. Participants were

exposed to a number of pairings of the symptoms and illnesses to determine whether their judgments about which symptoms predicted which illnesses altered over time. Over-selective responding would be shown if one element from the compound came to be rated as more predictive than the equally predictive second element from that compound; there being no difference in the degree to which these symptoms actually predicted the illness, it might be expected that both would be rated similarly to one another.

Method

Participants

Twenty-six participants (10 Males; 16 Females) were recruited from Swansea Psychology Department. Participants ranged from 19 to 56 years of age ($M = 28.54 \pm 9.77$). Participants received subject-pool credit for their participation. None of the participants reported any history of psychiatric or neurological illness. The study received approval from the Department of Psychology Ethics Committee.

Apparatus and Stimuli

All stimuli were presented, and responses recorded, through the experimental software package OpenSesame, version 0.27.3 *Frisky Freud* (Mathôt, Schreij, and Theeuwes, 2012). The six symptoms, A, B, C, D, E and F that participants rated were: ‘Bad Breath’, ‘Blurred Vision’, ‘Ear Ache’, ‘Nose Bleed’, ‘Skin Rash’, and ‘Stomach Ache’. All stimuli were presented in black letters, in Arial 24 font, in the center of one of four spherical turquoise discs that were presented in the corners of each slide on each trial. Three symptoms would be present on any one slide (leaving one disc containing no stimulus). Each slide was presented for 1000ms, after which a turquoise fixation cross appeared in the center of the screen for 1000ms, before being followed by an outcome, also presented for 1000ms.

There were three possible outcomes: one of two fictitious illnesses, or no illness. The two fictitious outcome illnesses were labelled 'Jetson's Syndrome', and 'Hartley's Disease', and these would appear in black letters, in Arial 32 font, in the center of a large turquoise disc, positioned in the center of the screen. This would occur for 12 trials of each block (i.e., 6 slides would each predict one of the fictitious illnesses). The 'No illness' outcome was denoted by a blank slide, which would occur on 12 trials of each block.

Upon completing a block of 24 trials, participants were presented with two successive ratings instruction screens, one for each illness. Each screen requested participants to provide a rating for each of the six symptoms in terms of how strongly it predicted the illness. The ratings were given on a scale ranging from 0 to 100 (where 0 = 'Not at all predictive'; and 100 = 'Completely predictive').

Procedure

All participants were tested individually in a small laboratory room. Participants were instructed to look at a computer monitor that provided the following test instructions:

"*In this experiment you will be shown a series of slides. Each slide will show four discs, some discs will contain the name of a symptom. Some of these slides will be followed by the name of illness, some will not. Your task is to learn which symptoms predict which illness. Please press the spacebar to begin the experiment.*"

Participants were then presented with the list of the symptoms, and one set of ratings, on a 0 to 100 scale, was requested for their judgment about how likely each symptom was to be associated with each illness. The order in which the symptoms were listed on the screen (top to bottom) was randomly determined for each screen, and the order of illnesses for which the symptoms were rated was randomly determined for each trial. Once participants had provided their ratings for each symptom for both illnesses, they were presented with a slide prompting them to proceed to the next block.

The participants were then exposed to 4 blocks of 24 trials each. The 6 symptom stimuli were designated to one of three sets for each participant: A+. BC+, and DEF-; where A+ represented the elemental stimulus predictive of one of the fictitious illnesses (e.g., Bad Breath = Jetson's Syndrome); BC+ represented the compound stimulus predictive of the alternative fictitious illness (e.g., Blurred Vision' and Stomach Ache = Hartley's Disease). The A+ and BC+ stimuli would never appear on the same slides together. The stimuli set defined as DEF- consisted of the non-predictive stimuli (D, E, and F), which were randomly presented with both the elemental stimulus (A+), and the compound stimulus (BC+), and which would also all appear together (without either the elemental stimulus or the compound stimuli) for 12 trials of each block, resulting in the no illness outcome (an example trial can be seen in Figure 1).

The slide order was randomized for each block of training. The actual stimuli components designated to each of the stimuli sets (e.g., A+ = Bad Breath) were the same within-participant (i.e., the same stimuli would reliably predict the same outcomes), but differed between-participant, the assignment of stimuli to be A, B, C, etc., was random for each participant.

After each block had been presented, the ratings instructions screens were presented for each illness. One set of ratings, on a 0 to 100 scale, was requested for each symptom with respect to each of the illnesses. The order in which the symptoms were listed on the screen (top to bottom) was randomly determined for each screen, and the order of illnesses for which the symptoms were rated was randomly determined for each trial. Once participants had provided their ratings for each symptom for both illnesses, they were presented with a slide prompting them to proceed to the next block. This was the case until they had completed the experiment.

Figure 1 about her

Results and Discussion

The pre-task mean ratings for the six symptoms with respect to Jetson's Syndrome were Blurred Vision = 25, Stomach Ache = 27, Bad Breath = 17, Ear Ache = 15, Nose Bleed = 24, Skin Rash = 27. The mean ratings for the six symptoms with respect to Hartley's Disease were Blurred Vision = 21, Stomach Ache = 22, Bad Breath = 15, Ear Ache = 19, Nose Bleed = 21, Skin Rash = 25. A two-factor repeated-measures analysis of variance (ANOVA) with symptom and disease as factors was conducted on these data, and revealed no statistically significant main effects or interaction, all Fs < 1, largest η^2_p = .014.

Figure 2 about here

Figure 2 shows the mean ratings for each stimulus across the four blocks of training. Stimulus A+ refers to the elemental stimulus, stimuli B+ and C+ denote the components of the compound (with B+ representing the component which received the higher rating at the first rating point), and stimuli D, E, and F signify the non-predictor stimuli. In order to obtain a mean non-predictor value, the ratings for the three non-predictor stimuli were summed and divided by 3. Examination of Figure 2 reveals that all predictive stimuli received increasingly higher ratings as training progressed. The ratings for the elemental stimulus (A) and the over-selected stimulus from the compound (B) increased to similar levels as one another. The under-selected stimulus (C) was rated at a similar levels as the non-predictor stimuli.

A two-factor repeated-measures ANOVA was performed on the data, with stimuli (A+, B+, C+, DEF-) and block as factors. In light of sphericity violations for the two factors

and interaction, smallest $\chi^2(5) = 13.99$, $p < .05$, the degrees of freedom were corrected using the Greenhouse-Geisser estimates. The ANOVA revealed statistically significant main effects of stimuli, F (2.37, 59.34) = 9.27, p. < 001, $\eta^2_p = .24$, 90% CI = [.10, .39],[1] and block, F (2.14, 53.44) = 12.10, p. < 001, $\eta^2_p = .31$, 90% CI = [.14, 45]. The interaction effect between these two factors was not significant, F (5.26, 131.57) = 1.69, $p > .10$, $\eta^2_p = .06$.

The ratings were collapsed across the 4 blocks, and analyzed using simple effects, which revealed that A+ was rated higher than the lower-rated component of the compound (C+), F (1, 25) = 15.17, $p < .01$, $\eta^2_p = .37$, 90% CI = [.12, .55]. The higher-rated component of the compound (B+) also received significantly higher ratings than the lower-rated component of the compound (C+), F (1, 25) = 15.2, $p < .01$, $\eta^2_p = .38$, 90% CI = [.12, .54]. The element (A+) did not significantly differ from the higher-rated component of the compound (B+), $F < 1$, $\eta^2_p = .00$.

Thus, an over-selectivity effect was noted for this study, with one of the elements of the compound being rated similarly to an element presented individually. Such findings have been noted in procedures related to associative learning paradigms (see Reed & Gibson, 2005), but not previously in such a diagnostic judgement task. Of course, selecting the higher-rated stimulus individually for each participant would be expected to produce such a result, although it should be remembered that this assignment was only made on the basis of the performance on block 1, and this would not necessarily mean the stimuli would be rated in this manner after each subsequent block.

1 Confidence intervals, set at 90%, were calculated using the MBESS R Package maintained by Kelley (2015).

Experiment 2

Experiment 2 utilized a between-subjects iteration of Experiment 1 in order to explore whether the findings were influenced by the multiple ratings given by participants regarding the stimuli, which is not usual in such studies of human causal judgments (e.g., Catena, Maldonado, and Cándido, 1998). If the findings in Experiment 1 were replicated when participants had completed differing levels of training, but had provided a rating only at the end of the training, this would appear to strengthen their generality.

Method

Participants and Apparatus

Twenty-four participants (6 males; 18 females) took part in the experiment, all of whom were recruited from a Psychology Department. Participants ranged from 18 to 62 years of age (M = 34.79 $\pm$ 13.27). Participants received subject-pool credit for their participation. The study received approval from the Department of Psychology Ethics Committee. None of the participants had exposure to the previous experiment. The apparatus and stimuli were identical to those outlined in Experiment 1, with the omission of rating screens at the end of each block.

Procedure

The procedure was as described in Experiment 1, with the exception of the point at which participants provided their ratings. Participants in Group 1 (n = 12), were exposed to the equivalent of 2 blocks of training in Experiment 1 (i.e., 48 trials), before being presented with the rating screens; whilst participants in Group 2 (n = 12), were exposed to the equivalent of 8 blocks of training (i.e., 192 trials), before being presented with the ratings screens. The previous rating points (presented after every 24th trial) were replaced with

screens informing participants to “press the spacebar to proceed”. If participants wished to take a short break, they were informed to do so during these stages.

Results and Discussion

The initial (pre-task) mean ratings for the six symptoms with respect to Jetson’s Syndrome were Blurred Vision = 20, Stomach Ache = 29, Bad Breath = 21, Ear Ache = 19, Nose Bleed = 20, Skin Rash = 23. The mean ratings for the six symptoms with respect to Hartley’s Disease were Blurred Vision = 23, Stomach Ache = 22, Bad Breath = 17, Ear Ache = 21, Nose Bleed = 26, Skin Rash = 23. A two-factor repeated-measures ANOVA (symptom x disease) was conducted on these data, and revealed no statistically significant main effects or interaction, all Fs < 1, η^2_p = .01.

Figure 3 about here

Figure 3 illustrates the mean stimulus ratings for the element (A); the components of the compound (BC), with B denoting the stimuli which received the higher rating; and the mean of the non-predictor stimuli (DEF) for both groups. Inspection of these data revealed that, Group 2 produced higher ratings for the predictive stimuli than Group 1. However, the ratings for the non-predictor stimuli decreased as the level of training increased. The rating given to stimulus B was similar that given to the elemental cue (A), but was higher than the rating given to the other component of the compound (C), for both groups.

A two-factor mixed-model ANOVA was performed on these data, with group (Group 1 and Group 2) as a between-subject factors, and stimuli (A; B; C; DEF) as the within-subjects factor. This analysis revealed statistically significant main effects of group, F (1, 22) = 8.43, p < .005, η^2_p = .27, 90% CI [.04, .47], and stimuli, F (3, 66) = 9.89, p < .001, η^2_p =

.31, 90% CI [.13, .41], and a significant interaction between these two factors, F (3, 66) = 6.66, $p < .001$, $\eta^2_p = .23$, 90% CI [.07, .34]. Simple effect analyses revealed that the element (A) was rated significantly higher than the lower-rated component of the compound (C), for both Group 1, F (1, 22) = 3.45, $p < .05$, $\eta^2_p = .13$, 90% CI [.00, .34] and Group 2, F (1, 22) = 3.03, $p < .05$, $\eta^2_p = .12$, 90% CI [.00, .32]. The difference between the higher-rated component of the compound (B) and the lower rated component (C) of the compound approached significance for Group 1, F (1, 22) = 2.50, $p < .07$, $\eta^2_p = .09$, 90% CI [.00, .30], and was significant for Group 2, F (1, 22) = 3.40, $p < .05$, $\eta^2_p = .03$, 90% CI [00, .34]. The difference between the element (A) and the higher-rated component of the compound (B) was not significant in either group, $F < 1$, $\eta^2_p = .007$. Thus, these results replicated the unilateral overshadowing effect observed in Experiment 1.

General Discussion

The present studies explored the existence of over-selectivity effects in a variant of a casual judgement task (e.g. Price and Yates, 1993; Van-Hamme and Wasserman, 1994) for typically-developing healthy participants. An over-selectivity effect was observed in both experiments, and is novel for a human causality judgment study. Typically, these effects have been demonstrated under procedures which are heavily based on basic conditioning studies (e.g. Reynolds and Reed, 2011). It has, therefore, not been clear whether these basic procedures could inform understanding of the processes occurring in real world decision making. However, observing over-selectivity in the current experiments with healthy participants, further demonstrates the generality of the effect and highlights the importance of understanding this phenomenon for everyday situations.

In the current context, it has been noted how these findings relate to medical decision making processes such as diagnosis. Diagnoses are often made on the basis of multiply-

presented symptoms, therefore it is clear to see how over-selective responding could be displayed in these situations and the potential implications of this effect in this context. Over-selecting a symptom as a cause of a particular illness (or allocating greater significance to a particular symptom) could result in misdiagnosis, or delay identification of a particular illness. It is often the case that the rapidity with which a correct diagnosis can be made is crucial in ensuring that the appropriate treatment is administered in a timely and cost effective manner (see Bonner, Monroe, Talley, Klasner, & Kimberlin, 2003). Thus, ensuring that the full range of symptoms are identified and that the predictive relationship between symptoms and a particular illness is understood is of crucial importance for both the patient's outcome and the efficient use of medical professionals' time and resources (see Fuat, Hungin & Murphy, 2003).

In the current study both symptoms of the compound were equally predictive of the same illness, thus attributing greater importance to one of these symptom's would have no impact on participants' ability to identify the correct illness (assuming they were able to accurately predict the illness on the basis of the other symptom they deemed to be of greater importance). This procedural arrangement was necessary to ensure there was some consistency between this study and others which have explored over-selectivity; if there were differential predictive relationships between the symptoms of the compound and the illness, one would expect a difference in the ratings these symptoms received.

Given this experimental procedure it is conceivable that, within this experiment, over-selectivity reflects optimal task performance as opposed to a cognitive deficiency. For example, the BC+ compound was always indicative of an illness. Therefore, to learn the relationship between this compound and the illness in the least cognitively taxing manner, it is only necessary to learn the relationship between one element of the compound and the illness. Employing this strategy would result in reduced cognitive load with no cost to accuracy. Of course, in the real world, however, attributing greater prominence to one

symptom at the expense of another, could be pivotal if the less selected symptom is indicative of another illness (e.g. B and C shared different contingencies with the illness), or in fact serves as evidence that the initial diagnosis was incorrect (e.g. the presence of the under selected symptom would rule out the initial diagnosis). This is a factor which would appear to be of interest for subsequent work in this area to consider now that the over-selectivity effect has been observed in these contexts (e.g. if symptoms of the compound were initially equally predictive but then become correlated at differential rates as predictive of other illnesses).

Given that the effect has now been obtained in the experimental procedures used in the set of experiments reported here, it might also be interesting to further extend these results into the area of medical-decision making, as an analogue for the impacts on this process of high levels of concurrent demands on medical staff (se Graber, 2007, for a review). In particular, it may be a useful procedure through which to further examine the impact of a range of factors on the degree to which some symptoms can come to control a diagnosis at the expense of others (Maserejian et al., 2009; Resnick et al., 1996). Understanding the factors which influence this process will enable insight in the mechanisms which mediate over-selective responding, and thus potentially provide a framework to consider how over-selectivity effects could be mitigated.

In the current case, it was noted that one stimulus from the compound was rated as predictive for the illness as a single symptom, but that the other symptom from the compound was not regarded as predictive. The symptom that was picked as being predictive varied from individual to individual. One way to interpret these effects which may provide insight into factors underpinning it, is through associative accounts of learning. Interestingly, models of learning, such as the Rescorla-Wagner (1972) model, would not suggest the result obtained in the experiments reported here. This model suggests that cues from a compound stimulus compete for a finite amount of associative strength, and postulates that such cue-competition effects such as over-selectivity will be reciprocal in nature; that is, both components are

predicted to accrue less learning (or elicit a lower rating) than elements trained in isolation (e.g. James & Wagner, 1980). However, the current findings suggest that the nature of the effect is unilateral rather than reciprocal, as the higher rated element of the compound received comparable ratings to the element trained in isolation.

It should be noted though, that this inference derived from the model assumes that cues will differ in salience at the outset (e.g., Le Pelley, Beesley, and Griffiths, 2014; Mackintosh, 1976), but will have no difference in their initial response strength (e.g. overshadowing). Studies on over-selectivity, however, often employ stimuli for which there may be difference in individual preference at the outset. In the current experiments, it is impossible to judge a priori the salience of the stimulus used, although some estimation can be obtained by the pre-task judgments, which showed no difference overall in the degree to which the stimuli were judged to be related to the symptoms. However, individuals may have come to the study with idiosyncratic views about the relationship between symptoms and diseases, meaning that response strengths differed from the outset. Although, in some senses, this idiosyncratic responding is inherent in the notion of over-selective responding, it was possible in the current data to determine if the stimulus that was rated as most predictive before training was the same as that which was rated as most predictive from the BC compound after training.

For Experiment 1, 6/26 participants initially selected the stimulus that became B in the BC compound as the most likely to predict the illness to which it was associated. This number was 5/24 in Experiment 2. When comparing the stimuli that were to become B and C, 14/26 participants rated the B stimulus as more predictive of the illness subsequently associated with that symptom in Experiment 1. In Experiment 2, this number was 11/24. These data suggest that the initial views of the participants had little impact on their ratings once the experiment had started, and that some other features of the stimuli present on the initial learning trials were responsible for the over-selection.

In sum, the present results revealed that over-selectivity can be replicated in a diagnostic judgment procedure with healthy participants, with one component of the compound receiving greater ratings than the other. In addition, a unilateral effect was observed whereby one component of the compound received ratings comparable to a predictor trained in isolation of any other predictor stimuli. These results may serve to suggest a potential experimental procedure for the study of over-selection during diagnosis, and may also serve to have a disconfirmatory status with respect to simple versions of the Rescorla-Wagner learning model in this context.

References

Allen, K. D., & Fuqua, W. R. (1985). Eliminating selective stimulus control: A comparison of two procedures for teaching mentally retarded children to respond to compound stimuli. Journal of Experimental Child Psychology, 39, 55-71.

Bonner, A. B., Monroe, K. W., Talley, L. I., Klasner, A. E., & Kimberlin, D. W. (2003). Impact of the rapid diagnosis of influenza on physician decision-making and patient management in the pediatric emergency department: results of a randomized, prospective, controlled trial. *Pediatrics*, *112*(2), 363-367.

Catena, A., Maldonado, A., & Cándido, A. (1998). The effect of frequency of judgement and the type of trials on covariation learning. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 481.

Chiang, H., & Carter, M. (2008). Spontaneity of communication in individuals with autism. *Journal of Autism and Developmental Disorders,* **38,** 693–705.

Dube, W. V., Balsamo, L. M., Fowler, T. R., Dickson, C. A., Lombard, K. M. & Tomanari, G. Y. (2006). Observing behavior topography in delayed matching to multiple samples. *The Psychological Record,* **56,** 233–244.

Dube, W. V., & McIlvane, W. J. (1999). Reduction of stimulus overselectivity with non-verbal differential observing responses. *Journal of Applied Behaviour Analysis,* **32,** 25–33.

Fuat, A., Hungin, A. P. S., & Murphy, J. J. (2003). Barriers to accurate diagnosis and effective management of heart failure in primary care: qualitative study. *Bmj*, *326*(7382), 196.

Graber, M. L. (2007). Diagnostic errors in medicine: What do doctors and umpires have in common. *Morbidity & mortality*, *2*, 1-6.

Huguenin, N. H. (1997). Employing computer technology to assess visual attention in young children and adolescents with severe mental retardation. *Journal of Experimental Child Psychology*, **65**, 141-170.

James, J. H., & Wagner, A. R. (1980). One-Trial Overshadowing: Evidence of Distributive Processing. Journal of Experimental Psychology: *Animal Behavior Processes*, 6, 188 – 205.

Keinan, G. (1987). Decision making under stress: scanning of alternatives under controllable and uncontrollable threats. Journal of personality and social psychology, 52(3), 639.

Kelley, K. (2015). MBESS (Version 4.0.0 and higher) [computer software and manual], Accessible from http://cran.r-project.org.

Kelly, M. P., Leader, G., & Reed, P. (2016). Factors producing over-selectivity in older individuals. AGE, 38(3), 1-10.

Koegel, R. L., Schreibman, L., Britten, K., & Laitinen, R. (1979). The effects of schedule of reinforcement on stimulus overselectivity in autistic children. *Journal of Autism and Developmental Disabilities,* **9,** 383-396.

Leader, G., Loughnane, A., McMoreland, C., & Reed, P. (2009). The effect of stimulus salience on over-selectivity. *Journal of Autism and Developmental Disorders*, *39*(2),

330-338.

Le Pelley, M. E., Beesley, T., & Griffiths, O. (2014). Relative salience versus relative validity: Cue salience influences blocking in human associative learning. Journal of Experimental Psychology: Animal Learning and Cognition, 40(1), 116

Lovaas, O. I., Koegel, R. L., & Schreibman, L. (1979). Stimulus overselectivity in Autism: A review of research. *Psychological Bulletin,* **86,** 1236–1254.

Lovaas, O. I., & Scheribman, L. (1971). Stimulus over selectivity of autistic children in a two stimulus situation. Behaviour, Research and Therapy, 9, 305-Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. Psychological Review, 82, 276-298.

Mackintosh, N.J. (1976). Overshadowing and stimulus intensity. Animal Learning and Behaviour, 4, 186-192.

Maserejian, N. N., Link, C. L., Lutfey, K. L., Marceau, L. D., & McKinlay, J. B. (2009). Disparities in physicians' interpretations of heart disease symptoms by patient gender: results of a video vignette factorial experiment. Journal of Women's Health, 18(10), 1661-1667.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. Behavior Research Methods, 44(2), 314-324. doi:10.3758/s13428-011-0168-7.

Price, P.C., & Yates, J. F. (1993). Judgmental overshadowing: Further evidence of cue interaction in contingency judgment. Memory & Cognition, 21, 561-572.

Reed, P. (2006). The effect of retention interval on stimulus over-selectivity using a matching to sample paradigm. *Journal of Autism and Developmental Disorders,* **36,** 1115–1121.

Reed, P., & Gibson, E. (2005). The effect of concurrent task load on stimulus over-selectivity. Journal of Autism and Developmental Disorders, 35(5), 601-614.

Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A. H., & Prokasy, W. F. (Eds.), Classical conditioning II: Current research and theory (pp. 64-99). New York: Appleton-Century-Crofts.

Resnick, N. M., Brandeis, G. H., Baumann, M. M., DuBeau, C. E., & Yalla, S. V. (1996). Misdiagnosis of urinary incontinence in nursing home women: prevalence and a proposed solution. Neurourology and urodynamics, 15(6), 599-618.

Reynolds, G., & Reed, P. (2011). The strength and generality of stimulus overselectivity in simultaneous discrimination procedures. *Learning and Motivation*, **42**, 113-122.

Stromer, R., McIlvane, W. J., Dube, W. V., & Mackay, H. A. (1993). Assessing control by elements of complex stimuli in delayed matching to samples. *Journal of the Experimental Analysis of Behaviour,* **59,** 83–102.

Van Hamme, L.J., & Wasserman, E.A. (1994). Cue competitions in causality judgments: The role of nonpresentation of compound stimulus elements . *Learning and Motivation*, 25, 127-151.

Wayland, S., & Taplin, J. E. (1985). Feature processing deficits following brain injury over selectivity in recognition memory for compound stimuli. *Brain and Cognition,* **4,** 338–355.

Figure Captions

Figure 1 - Example trial featuring the 'BC+" compound. Three symptoms would be present on any one slide. In this example two of the symptoms denoted 'B' and ''C' would predict the same illness, whilst the other symptom 'E' would be a distractor. Following the presentation of the symptoms (1000ms), a fixation cross would be presented for 1000m, before the illness would be presented for 1000ms.

Figure 2. Mean ratings for each stimulus across the four training blocks in Experiment 1. A+ = elemental stimulus presentation; B+ = component of the compound which received the higher rating at the first rating point; C+ = component of the compound which received the lower rating at the first rating block, and DEF- = non-predictor stimuli.

Figure 3. Mean ratings for each stimulus in Experiment 2 when the elements were defined by their initial ratings. A+, B+, C+ and DEF-, for Group 1, who were exposed to 48 trials, and Group 2, who were exposed to 192 trials. Standard error bars included.

Figure 1

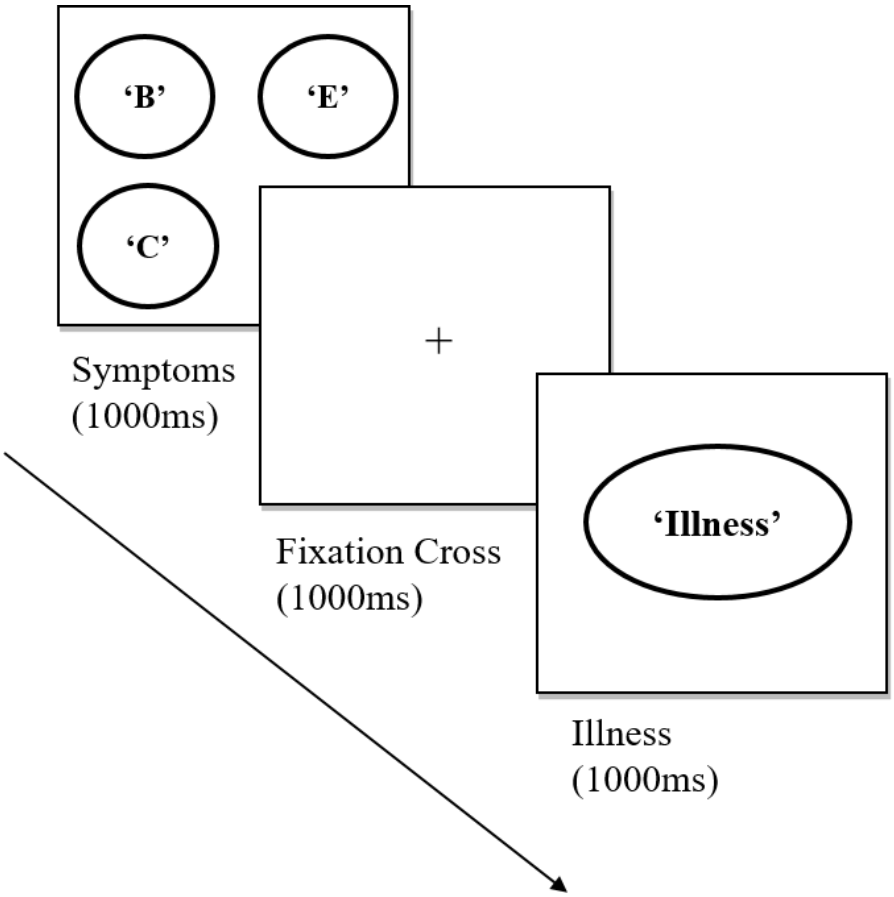

'B'
'E'
'C'
Symptoms
(1000ms)
+
Fixation Cross
(1000ms)
'Illness'
Illness
(1000ms)

Figure 2

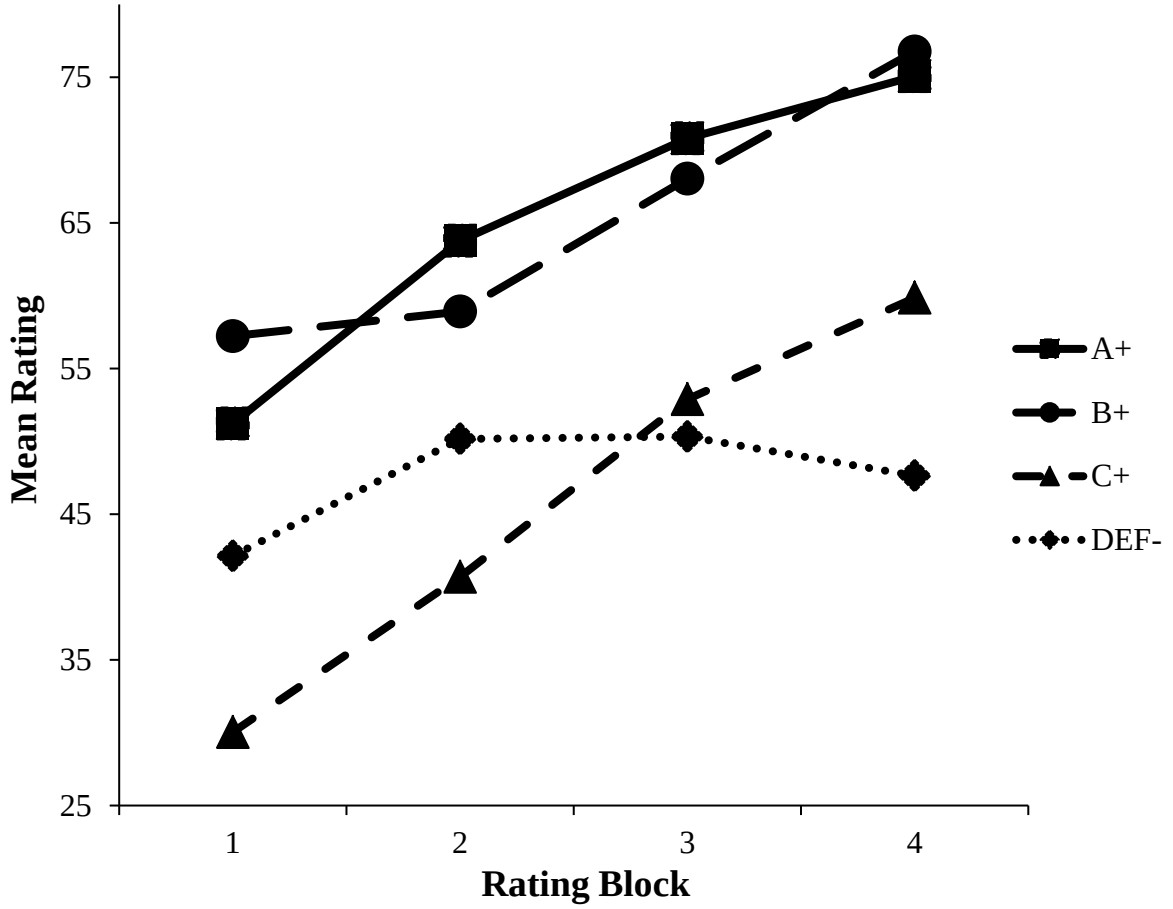
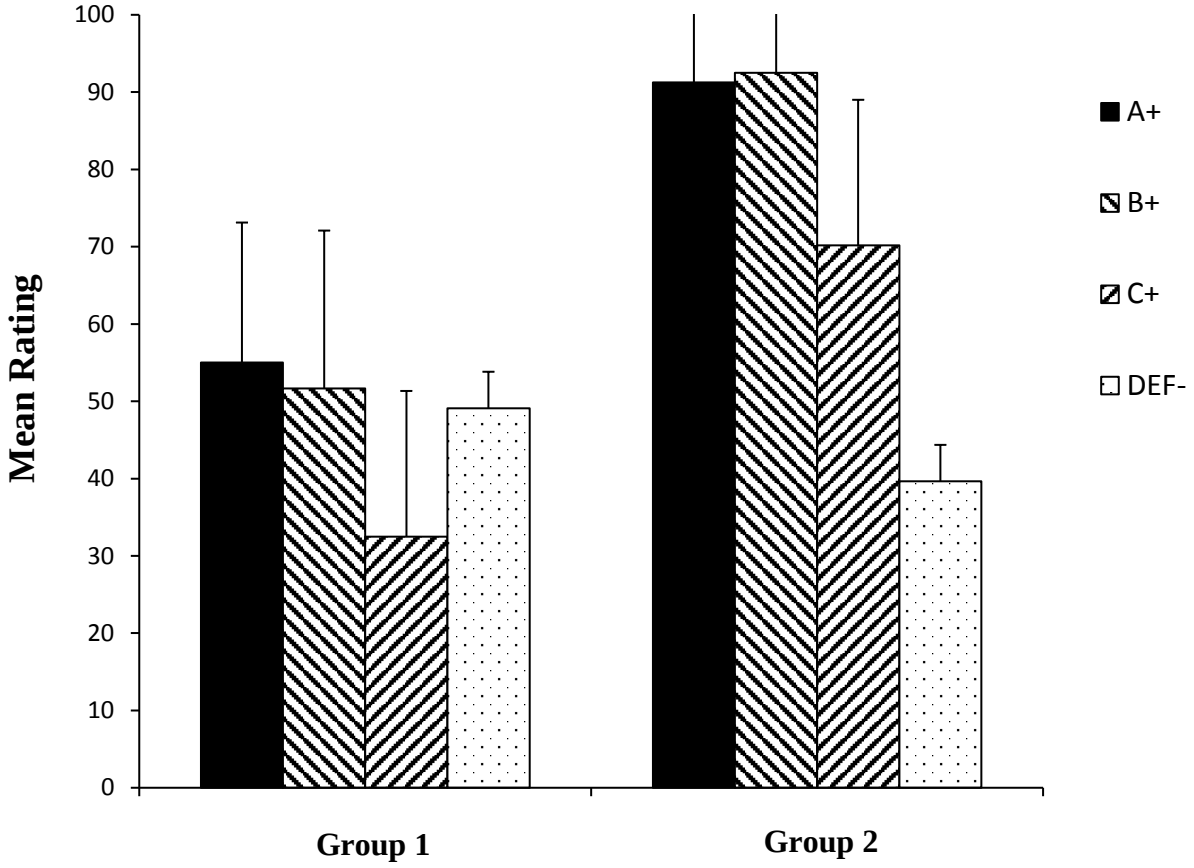


Figure 3


100
90
80
70
60
50
40
30
20
10
0
Mean Rating
A+
B+
C+
DEF-
Group 1
Group 2