

DOES THE TRUTH INTERFERE WITH OUR ABILITY TO DECEIVE?

Magda Osman<sup>1,2</sup>, Shelley Channon<sup>2</sup>, Sian Fitzpatrick<sup>2</sup>

<sup>1</sup>Department of Psychology, University of Surrey, Guildford, Surrey, GU2 7XH, UK

<sup>2</sup>Department of Cognitive, Perceptual and Brain Sciences, University College London, Gower Street, London, WC1E 6BT, UK

To whom correspondence should be addressed: [M.Osman@surrey.ac.uk](mailto:M.Osman@surrey.ac.uk)

Word count: 3803

## Abstract

Does the truth interfere with our ability to respond deceptively? We consider this question by examining the effects of task set (i.e. selecting truthful or untruthful responses), both by comparing two presentations of the same task, and through transfer to a different task. All participants carried out the task under instructions to respond correctly, and also to respond incorrectly (Experiment 1), or instructions to respond truthfully and also to respond deceptively (Experiment 2); order of instructions was counterbalanced. In Experiment 2, after completing the main task, participants also performed a Stroop task. The findings suggested that deceptive responses took longer overall, regardless of order of instructions. Moreover, experience of responding deceptively versus truthfully led to faster responding on a subsequent Stroop test. Whilst there may well be processes unique to deceptive responding, the evidence suggests that overcoming interference is a process shared by deceptive as well as non-deceptive tasks.

149 words

The study of deceptive communication has revealed that when making a deceptive response (e.g. saying “No” to the question “Do you smoke?” when in fact you do), people take longer than when responding truthfully (e.g., Vendemia et al., 2005; Walczyck et al., 2003, Walczyck et al., 2005). Slower processing is well documented in deception (see e.g. DePaulo, et al, 2003). For instance, in generating a lie, people typically remain close to the truth and omit or distort critical details (e.g., DePaulo et al., 2003). It seems that few people take the risk of generating a new account from scratch; and if they do, this is often aided by recruiting from memory other relevant experiences that can replace the truth (e.g., Johnson & Raye, 1981), and which are often highly associated with it (e.g. Sporer, 2004). The truth is thought to provide a marker for the credibility of the constructed lie, helping the deceiver to attenuate their guilt by rationalizing that their deceptive communication is “virtually true” (e.g., Bond & DePaulo, 2006).

Some (e.g., Jokinen, et al., 2006; Vendemia et al., 2005; Walczyck et al., 2003, Walczyck et al., 2005) claim that for deceptive responding, the presence of the truth is disadvantageous. Walczyck et al (2005) proposed that relevant truthful memories triggered when to-be-lied-to questions are presented may interfere with deceptive responding. They found that distance in time (remote or recent memories) influenced the degree of memory activation. The accessibility of true memories may therefore have a direct interfering influence on the speed of deceptive responding. Extra controlled processes are thus needed to prevent the truth from interfering, whether through downgrading or suppression of the truthful response and/or increased activation of the deceptive response through a selective attention mechanism (see e.g., Egner & Hirsch, 2005; Gorfein and MacLeod, 2007; Spence et al., 2004; e.g., Vendemia et al., 2005). In support, recent neuro-imaging evidence has revealed that activations when making deceptive responses are similar to those reported in studies examining high loading of executive functions (e.g., Langleben et al., 2005; Nunez et al., 2005).

Thus far, it still remains unclear what task characteristics influence any interfering effects of the truth in responding deceptively. To investigate this we consider the question from two perspectives – the degree of distance of the deceptive response from the truth, and the task set (namely, selecting truthful or untruthful responses); the latter is examined at two levels: within blocks of the same task, and through transfer to a different task.

We consider firstly whether selecting response options that are closely semantically related to the truth interferes more than selecting those distantly related to the truth, based on Walczyk et al's (2005) model. Using similar procedures to those used in forced-choice deception studies (Vendemia et al., 2005; Walczyk et al., 2003, Walczyk et al., 2005), we examine the generality of the interference effects by exploring this under both neutral (correct/incorrect) instructions (Experiment 1) and deception-based (true/lie) instructions (Experiment 2).

Secondly, we investigate whether experience of developing a task set to respond truthfully hinders or facilitates responding deceptively, under both neutral and deceptive instructions. Development of a task set involves establishing a pattern of responding (Mayr & Kliegl, 2000), and switching to an alternative task set involves overcoming interference from any previous set (e.g., Rogers & Monsell, 1995). Thirdly, we investigate another possible way that a task set may develop, by asking whether overcoming interference from an unwanted response is a general process found both in deceptive and non-deceptive tasks, sharing similar processing requirements to tasks such as the Stroop test (1935). This typically involves presenting color-word stimuli in conflicting ink colors (e.g. 'blue' in red ink) (incongruent trials), or in consistent ink colours (e.g. 'blue' in blue ink) (congruent trials). Typically, participants are slower to identify either the word or ink color when there is a mismatch between them. If making deceptive responses represents a form of incongruency, similar processes may be involved in overcoming interference in both deception-based forced-choice responding and in a Stroop task. We might then expect task-specific training gains, such that greater experience of deceptive responding should facilitate faster Stroop performance.

## Experiment 1

Experiment 1 examined the first hypothesis, that the presence of response alternatives that are semantically proximal to versus distant from the correct (truthful) response may produce greater interference effects, and if so, that these effects may be greater when making incorrect versus correct responses. We also examine the second hypothesis relating to within-task set by seeing whether experience of correct responding interferes with subsequent incorrect responding.



or right) was randomized across questions. The order of presentation of the 8 questions was pseudo-randomized for each scenario, so that no pair of related questions appeared consecutively.

All participants performed the 10-scenario task twice, once under instructions to give correct answers, and once under instructions to give incorrect answers. Participants were randomly assigned to one of two Instructions conditions: either correct-incorrect (respond correctly throughout the first task presentation, then incorrectly throughout the second task presentation) ( $n = 41$ ) or incorrect-correct (respond incorrectly first then correctly second) ( $n = 38$ ). Thus, when responding correctly, participants were required to respond selectively to the correct option and ignore the incorrect option (either proximal or distal), and when responding incorrectly they were required to respond only to the incorrect option (either proximal or distal) and ignore the correct option.

## Results and Discussion

Mean reaction times (RTs in msec) for distal and proximal response selections by Instructions and Order are presented in Figure 1. Errors in response selections were infrequent, and statistical analyses revealed no significant difference between failing to select distal (0.63%) and proximal (1.64%) response alternatives, and selecting the distal (2.15%) and proximal (3.79%) distractors. All analyses that follow are based on responses without errors, using logarithmically transformed data.

INSERT FIGURE 1 ABOUT HERE

### **Semantic distance**

Consistent with previous evidence (Vendemia et al., 2005; Walczyk et al., 2003, Walczyk et al., 2005), there was a significant main effect of Truth,  $F_{(1,77)} = 44.97$ ;  $P < 0.0005$ , partial  $\eta^2 = .37$ , such that responding correctly was faster than responding incorrectly. In order to examine any effect of semantic distance from the correct answer (truth) on incorrect responding, a 2x2x2 ANOVA was conducted with Semantic Distance (Distal, Proximal) and Truth (Correct, Incorrect) as within-group factors, and

Instructions (Correct – Incorrect, Incorrect – Correct) as a between-group factor. There was a main effect of Semantic Distance,  $F_{(1,77)} = 39.25$ ;  $P < 0.0005$ , partial  $\eta^2 = .34$ , showing that the presence of semantically proximal rather than distal information slowed responses. However, since there were no significant interactions involving Semantic Distance, participants were slowed whether they were ignoring the proximal response alternative (when instructed to respond correctly) or selecting the proximal response alternative (when instructed to respond incorrectly).

### **Within-task task set**

The effects of performing the task twice were examined with respect to the effects of prior experience of correct (truthful) responding on performance. There was a significant main effect of Instructions,  $F_{(1,77)} = 4.33$ ;  $P < .041$ , partial  $\eta^2 = .05$ , showing that those in the Correct-Incorrect condition responded faster overall than those in the Incorrect-Correct condition. There was also a significant Truth x Instructions interaction,  $F_{(1,77)} = 224.88$ ;  $P < 0.0005$ , partial  $\eta^2 = .75$ .

It can be seen from Figure 1 that both groups speeded up from the first to the second completion of the task, and that this effect was larger for the group switching from incorrect to correct responding. T-test comparison of incorrect responses (averaged across Semantic Distance) for the Correct-Incorrect group versus the Incorrect-Correct group showed that these were slower for the latter participants,  $t_{(77)} = 7.84$ ;  $P < 0.0005$ , Cohen's  $d = 1.85$ , suggesting that acquiring a task set for responding truthfully facilitated rather than interfered with subsequent incorrect responding, presumably reflecting some general effect of practice on the task.

## **Experiment 2**

Experiment 2 was designed to replicate and extend Experiment 1 by using deception-based instructions instead of neutral instructions, so that participants were told to select the 'lie' rather than the 'incorrect' response, and to select the 'true' rather than the 'correct' response. In addition, Experiment 1 suggested that prior practice at the task was sufficient to facilitate performance. To examine the second

hypothesis further by distinguishing between the effects of practice versus a specific task set, Experiment 2 included two additional conditions in which participants were instructed to lie or respond truthfully for both the first and second presentations of the task. Finally, it also explored the third hypothesis, the potential relationship between responding deceptively and any effects of transfer in overcoming later interference. To achieve this we included a version of the Stroop task at the end of the deception-based task. It was hypothesized that if there are shared processes involved in the two tasks, then experience of deceptive responding should facilitate performance on the Stroop task, particularly on incongruent trials.

## Methods

### *Participants*

Sixty four (50 females) UCL first-year undergraduate students (mean age = 20, s.d. 2.65) took part in the study.

### *Design & Procedure*

With the following three exceptions, in all other respects Experiments 1 and 2 were identical. Firstly, the instructions directed participants to respond truthfully or to lie rather than to respond with the correct or incorrect answer (True – Lie or Lie – True). Secondly, two additional practice conditions were added in which instructions for the first and second task completions were the same, involving either only truthful (True – True) or deceptive responses (Lie – Lie). Participants were randomly assigned to one of these four conditions: True – True, Lie – True, True – Lie, Lie – Lie, with 16 participants in each condition.

Thirdly, after completing the task twice, all participants carried out a Stroop task. A total of 40 congruent (i.e. ink and word compatible) and 40 incongruent (i.e. ink and word incompatible) trials were presented. Half the trials instructed participants to name the color of the ink, regardless of the word itself, and the remainder instructed them to name the word, regardless of the ink color. Order of presentation of word versus ink trials was counterbalanced across participants.

## Results and Discussion

As for Experiment 1, the frequency of errors for the main deception task was low, and analyses revealed no significant difference between responding incorrectly by failing to select distal (2.04%) and proximal (2.34%) response alternatives and by incorrectly selecting the distal (1.87%) and proximal (2.96%) distracters. The following analyses are based on correct RT responses, using logarithmically transformed data.

INSERT TABLE 1 ABOUT HERE

### **Semantic distance**

A 2x2x2x2 ANOVA was carried out using the data from all 4 conditions in Experiment 2, with Semantic Distance (Distal, Proximal) and Repetition (First or Second Task Completion) as within-group factors, and Instructions (True or Lie on first task completion) and Second Instructions (True or Lie on second task completion) as between-group factors (see Table 1). There was a significant main effect of Semantic Distance,  $F_{(1, 60)} = 46.67$ ;  $P < 0.0005$ , partial  $\eta^2 = .44$ , consistent with Experiment 1, showing that the presence of semantically proximal rather than distal information slowed responses. However, as in Experiment 1, there were no significant interactions involving Semantic Distance (see Figure 2).

INSERT FIGURE 2 ABOUT HERE

### **Within-task task set**

The ANOVA also showed a significant main effect of Repetition,  $F_{(1, 60)} = 124.53$ ;  $P < 0.0005$ , partial  $\eta^2 = .68$ , suggesting that overall, decisions were faster on the second completion of the task than the first. Experiment 1 suggested that incorrect responding was facilitated by task practice, since those who previously completed the task under correct instructions were faster than those with no prior task practice. However, this experiment could not differentiate between general task practice and any specific task set. To address this, we compared deceptive performance for those who previously completed the task under

instructions to tell the truth (general task practice and opposite task set) with those who previously completed it under instructions to lie (general task practice and same task set) (see Figure 2). Although the Lie-Lie group were significantly slower than the True-Lie group on the first completion of the task,  $t_{(30)} = 2.16$ ;  $P < 0.039$ , Cohen's  $d = .76$  (i.e. lying took longer than telling the truth), there was no significant difference on the second task completion in speed of deceptive responding between these two groups, suggesting that performance was affected primarily by general task practice rather than a specific task set.

### ***Effects of task set involving deceptive experience on the Stroop task***

To what extent does experience of deceptive responding influence responding on a subsequent interference task? The Stroop task was performed after the story task to examine the influence of prior practice in overcoming interference. Mean RTs for this are presented in Table 1. A 2x2x2x2 ANOVA with Congruency (Congruent, Incongruent trials) and InkWord (Respond-to-Ink-Color or Respond-to-Word-Meaning) as within-group factors and Instructions (True or Lie on first completion of story task) and Second Instructions (True or Lie on second completion of story task) as between-group factors.

The Stroop data showed the expected significant effect of Congruency,  $F_{(1,60)} = 66.36$ ;  $P < 0.0005$ , partial  $\eta^2 = .53$ , since responses to congruent trials were faster than to incongruent trials, and the expected significant effect of InkWord,  $F_{(1,60)} = 40.44$ ;  $P < 0.0005$ , partial  $\eta^2 = .40$ , since responding to the dominant word meaning was faster than responding to the ink color. There was also a significant InkWord x Congruency interaction,  $F_{(1,60)} = 16.06$ ;  $P < 0.0005$ , partial  $\eta^2 = .21$ , and no other interactions involving InkWord; a post-hoc paired t-test showed that the difference between incongruent and congruent trials was greater for Respond-to-Word than Respond-to-Ink-Color instructions,  $t_{(63)} = 4.03$ ;  $P < 0.0005$ , Cohen's  $d = .72$ .

With respect to the effects of prior deceptive experience on Stroop responding, there were no significant main effects of Instructions or Second Instructions nor interaction between them, and no interactions with Inkword; there was a significant Congruency x Instructions x Second Instructions interaction,  $F_{(1,60)} = 4.71$ ;  $P = 0.034$ , partial  $\eta^2 = .07$ . Mean scores showed that as predicted, experience of

responding deceptively facilitated Stroop performance (see Figure 3). Direct post-hoc comparison of the two practice groups combining across ink color and word trials showed that incongruent (but not congruent) responses on the Stroop task were significantly faster for those with the greatest prior experience of deceptive responding (Lie-Lie group) versus those with no prior experience of deceptive responding (True-True group),  $t_{(30)} = 2.15$ ;  $P < 0.040$ , Cohen's  $d = .76$ .

INSERT FIGURE 3 ABOUT HERE

## General Discussion

The present study was designed to examine the way in which the truth interferes with responding deceptively. Based on Walczyk et al's (2005) proposal that the activation of recent truthful memories interferes more than remote memories with the speed of deceptive responding, we considered a related issue, namely whether selecting response options that are closely semantically related to the truth interferes more than selecting those distantly related to the truth. Our findings did not support the notion of a specific interfering effect of truthful information on deceptive responding. Instead, the results showed semantic distance effects independent of task instruction, since both truthful and incorrect/deceptive responses were slower when the incorrect response alternative was closer to the truth. However, one consideration is that in Walczyk et al's (2003) and Walczyk et al's (2005) forced-choice studies, the items were self-generated and personal to the individual, whereas in the present study they were not. It may be the case that the relationship between semantic distance effects and deceptive responding is enhanced in highly personalized situations. One way in which this could be addressed is to use a similar task design to the one developed in the present study, but to include scenarios that are personal to the individual, and in which they generate the proximal and distal options.

Is it crucial that participants are intending to deceive, or is the critical factor that they are instructed to choose an incorrect answer? A similar pattern of responding was found across two experiments under both neutral (correct/incorrect) and deception-based (true/lie) instructions. It is of course possible that

participants do not perceive themselves to be actively engaged in deception unless they generate their own responses, rather than select from pre-determined ones as in the present study. However, face-to-face lying is also associated with slowed responding (for discussion see DePaulo, et al., 2003), and a number of previous studies in the deception literature have also used computer-based choice reaction tests in order to examine deceptive responding (e.g. Nunez et al, 2005; Spence et al, 2004; Vendemia et al, 2005; Walczyk et al, 2003; Walczyk et al, 2005). We contend that interference effects may play an important role in the selection of deceptive responses, and that these are unlikely to be specific to deception tasks.

Interference effects were also examined in the present study in relation to task set, both within blocks of the same task, and through transfer to a different task. Within blocks of the same task, the findings suggest that developing a task set for responding truthfully (by performing the task first under instructions to select truthful responses) does not in itself interfere with responding deceptively; indeed, there may even be a facilitatory role. Faster performance on the second presentation of the task was found across instructional orders on the two experiments. The most parsimonious explanation for this is an effect of task practice.

With respect to transfer to a different task, both the deception and Stroop tasks involved overcoming the interfering effects of task-irrelevant items. We postulated that if processes involved in overcoming interference contributed to slower deceptive responding, then experience of this type of response in the deception task should facilitate performance on an unrelated task that requires similar processing. The clearest support for this was provided by comparison of the two conditions in Experiment 2 that gave either no experience (True-True) or maximal experience (Lie-Lie) of deceptive responding. Practice in responding deceptively rather than truthfully was found to enhance performance on an unrelated Stroop task. This facilitation effect suggests that development of a general task set to respond deceptively may require selective attention to items relevant to the task at hand, whilst overcoming interference from task-irrelevant items. Recent work by Dahlin, Stigsdotter Neely, Larsson, Bickman, & Nyberg (2008) suggested that transfer of learning effects can be seen on working memory tasks, but only if they are mediated by the same neural system. A similar approach could be taken to elucidate more fully the general

underlying processes involved in deceptive responding, by comparing any transfer effects for subsequent tasks thought to share the same underlying processes, as well as for other tasks which do not.

## Acknowledgments

The support of the ESRC Research Centre for Economic Learning and Human Evolution is gratefully acknowledged.

## References

- Bond, C. F., & DePaulo, B. (2006). Accuracy of Deceptive Judgments. *Personality and Social Psychology Review*, *10*, 214-234.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to Deception. *Psychological Bulletin*, *129*, 74–118.
- Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*(12), 1784–1790.
- Gorfein, D.S., & MacLeod, C.M. (2007). *Inhibition in cognition*. Washington, DC: American Psychological Association.
- Johnson, M. K., & Raye, C. L. (1981). Reality Monitoring. *Psychological Review*, *88*, 67-85.7,
- Jokinen, A., Santtila, P., Ravaja, N., & Puttonen, S. (2006). Salience of guilty knowledge test items affects accuracy in realistic mock crimes. *International Journal of Psychophysiology*, *62*, 175-184.
- Langleben, D. D., Loughhead, J. W., Bilker, W. B., Ruparel, K., Childress, A. R., Busch, S. I., & Gur, R. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping*, *26*, 262–272.
- Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1124–1140.
- Nunez, J. M., Casey, B. J., Egner, T., Hare, T., & Hirsch, J. (2005). Intentional false responding shares neural substrates with response conflict and cognitive control. *Neuroimage*, *25*, 267–277.
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207–231.
- Spence, S. A., Hunter, M. A., Farrow, F. D., Green, R. D., Leung, D. H., Hughes, C. J., & Ganesan, V. (2004). *A cognitive neurobiological account of deception: Evidence from functional neuroimaging*. *Philosophical Transactions of the Royal Society of London B*, *359*, 1755–1762.
- Sporer, S. L. (2004). Reality monitoring and the detection of deception. In P.-A. Granhag and L. Strömwall

(Eds.), *Deception Detection in Forensic Contexts*. Cambridge: Cambridge University Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.

Vendemia, J. M., Buzan, R. F., & Green, E. P. (2005). Practice effects, workload, and reaction time in deception. *American Journal of Psychology*, 118, 413-429.

Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology*, 17, 755-774.

Walczyk, J. J., Schwartz, J. P., Clifton, R., Adams, B., Wei, M., & Zha, P. (2005). Lying person-to-person about life events: A cognitive framework for lie detection. *Personnel Psychology*, 58, 141-170.

Table 1. Mean scores and standard deviations for Experiments 1 and 2.

	<b>Mean</b>	<b>(SD)</b>	<b>Mean</b>	<b>(SD)</b>	<b>Mean</b>	<b>(SD)</b>	<b>Mean</b>	<b>(SD)</b>
Story Task	Correct/Truthful Alternative				Incorrect/Deceptive Alternative			
Semantic Distance	Proximal		Distal		Proximal		Distal	
Experiment 1								
Instructions								
<b>Correct-Incorrect</b>	1410.56	(351.26)	1344.11	(317.61)	1218.85	(275.93)	1146.40	(236.72)
<b>Incorrect-Correct</b>	1180.67	(250.88)	1098.33	(211.28)	1737.94	(367.62)	1627.48	(358.49)
Experiment 2								
Instructions								
<b>True-Lie</b>	1442.57	(422.15)	1311.07	(362.67)	1341.44	(344.95)	1259.56	(332.34)
<b>Lie-True</b>	1354.22	(515.19)	1217.98	(400.43)	2030.19	(536.47)	1707.90	(396.16)
<b>True-True</b>	1621.03	(428.29)	1372.07	(291.51)	1072.68	(195.95)	1011.99	(138.95)
<b>Lie-Lie</b>	1764.35	(513.38)	1650.28	(528.06)	1414.94	(387.36)	1288.23	(306.30)
Experiment 2								
Stroop Task	Congruent				Incongruent			
Instructions	Ink		Word		Ink		Word	
<b>True-Lie</b>	1105.58	(573.46)	804.00	(390.05)	1205.68	(465.49)	1093.64	(625.79)
<b>Lie-True</b>	1094.48	(411.30)	898.79	(317.64)	1158.89	(341.91)	1159.92	(435.14)
<b>True-True</b>	1218.47	(415.19)	985.71	(405.51)	1318.50	(387.35)	1141.60	(438.81)
<b>Lie-Lie</b>	1031.13	(300.52)	793.44	(207.83)	1040.75	(281.23)	942.17	(184.40)

Production Number R527B:

Does the truth interfere?

Figure Captions

*Figure 1:* Mean RTs and standard errors of mean by Instructions and Response Distance in Experiment 1

*Figure 2:* Mean RTs and standard errors of mean by Instructions and Response Distance in Experiment 2

*Figure 3:* Mean RTs and standard errors of mean for Congruent and Incongruent Stroop trials averaged across ink and word stimuli, by Task Instruction in Experiment 2.

Figure 1:

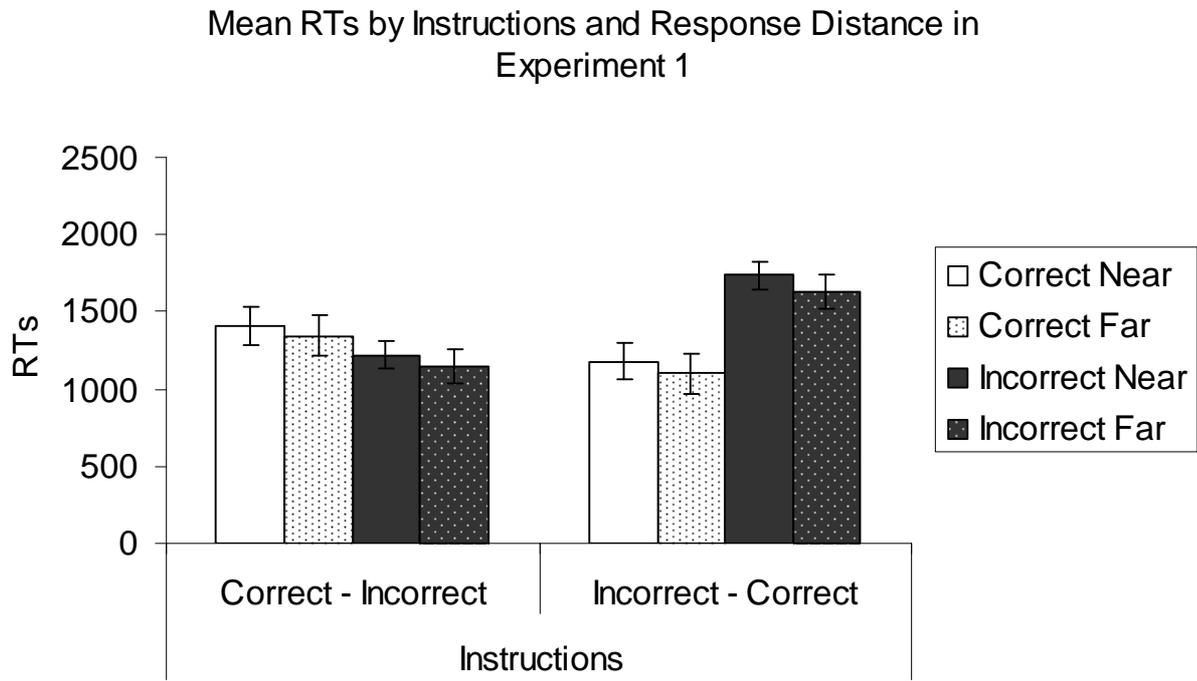


Figure 2:

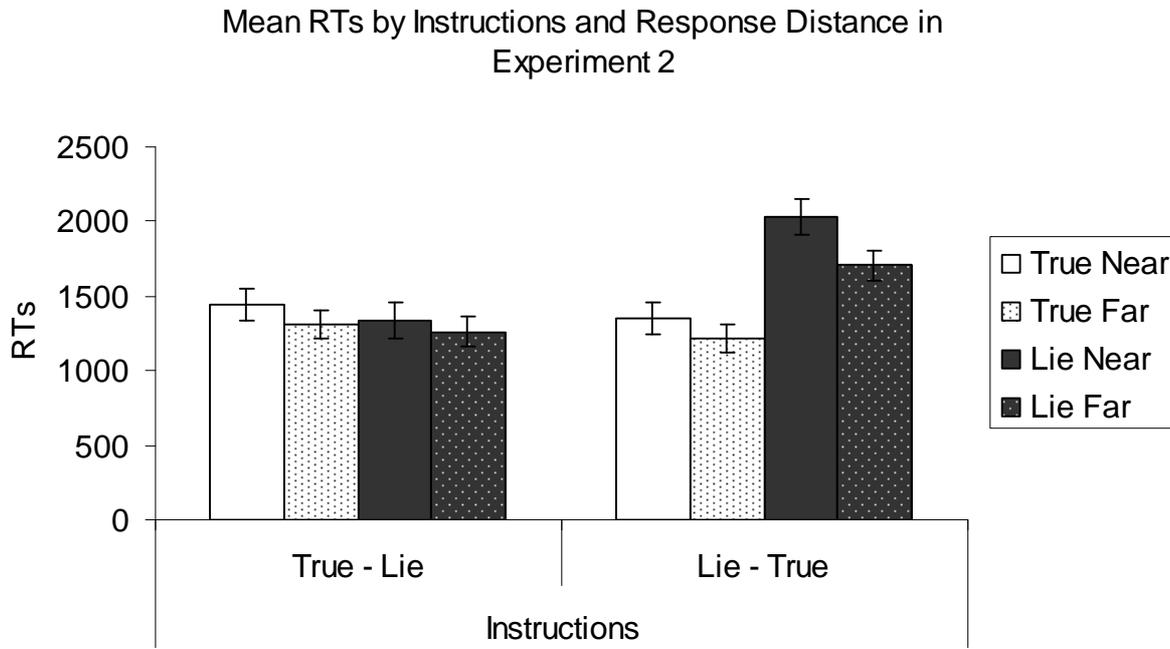


Figure 3:

