A Model of the “Guilty Knowledge Effect:” Dual Processes in Recognition

Travis Seymour
Colleen Seifert

University of Michigan

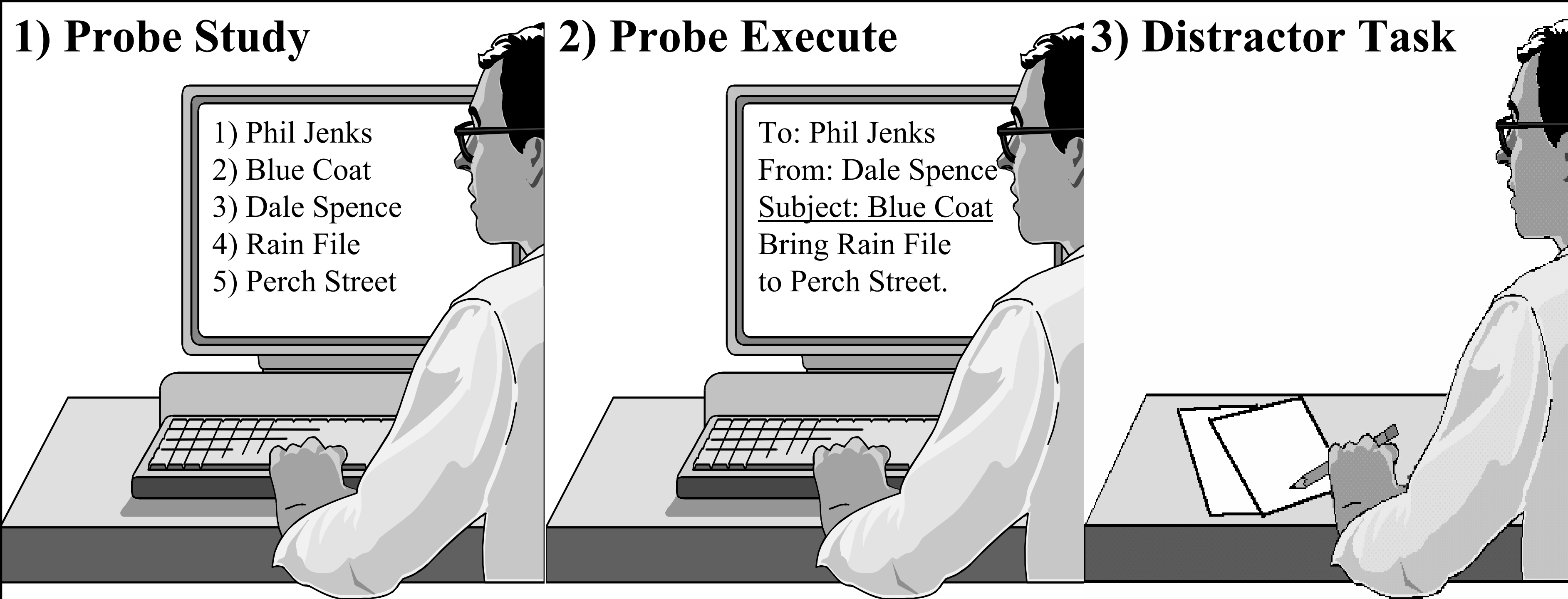
Introduction

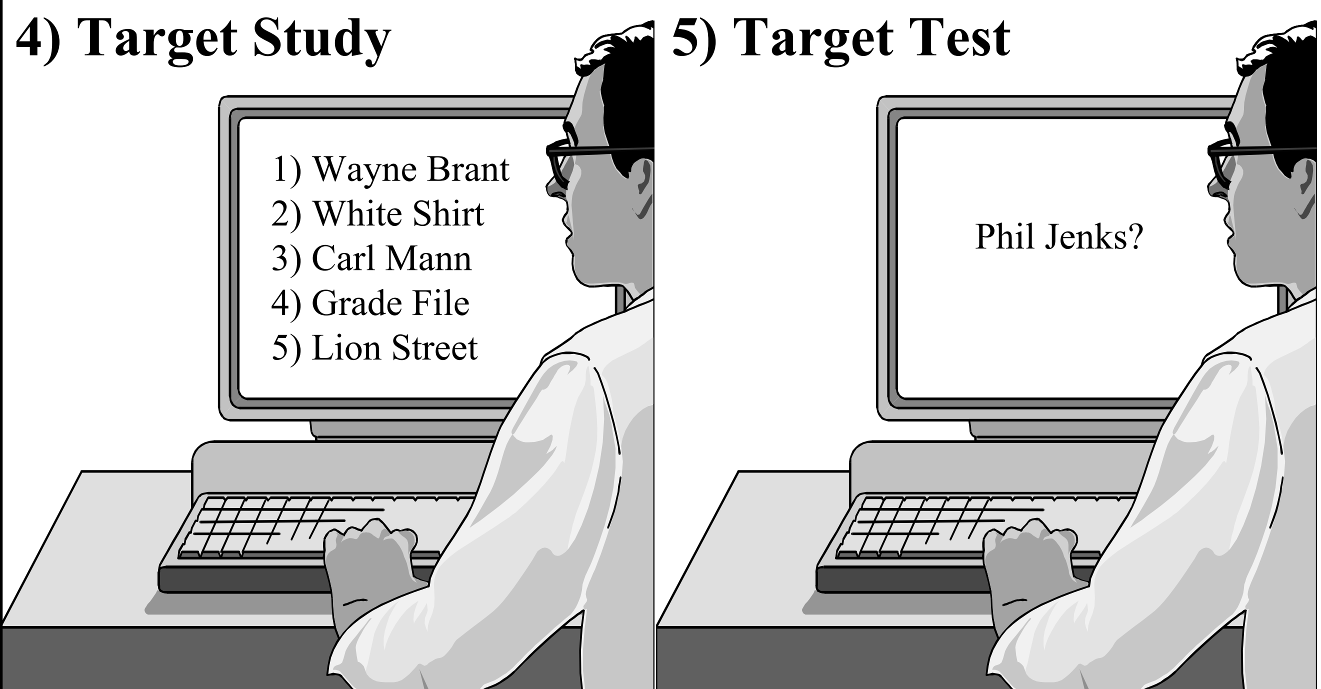
Even when a person wishes to misrepresent the contents of their memory, their memory may still be accessible. One method of determining whether particular knowledge is available to an individual is the polygraph-based Guilty Knowledge Test (GKT) (e.g. Lykken, 1981). With this test, a suspect's attempt to appear ignorant of facts actually familiar to them is revealed by physiological measures such as heart rate and galvanic skin response. While results from such tests are admissible in many courts, when both sides consent, there are numerous studies questioning their reliability and validity (e.g., Bashore & Rapp, 1993; Furedy & Heslegrave, 1988).

As an alternative, Farwell & Donchin (1991) developed a measure of Guilty Knowledge based on knowledge activation in memory rather than physiological response. By examining the P300 component of a suspect's evoked related potentials (ERP) when presented with crime-related probes (versus neutral stimuli), Farwell & Donchin (1991) were able to accurately classify 90% of "guilty" and 85% of "innocent" participants. Because ERPs are considered rather difficult to manipulate and do not necessitate interpretation (as polygraphs do), the authors posit this method as a superior alternative.

The Guilty Knowledge Effect

Considering the cost of training, administration and analysis associated with ERPs, Seymour, Mosmann and Seifert (1997) developed a simpler variation of Farwell & Donchin's (1991) methodology using response times (RT) instead of ERPs. Despite speculation that a Guilty Knowledge Test based on RTs would be too manipulable to be effective, Seymour et al. found that RTs, as well as response accuracy, yielded a high classification accuracy (93% "guilty" and 100% "innocent"). Because of a response deadline too fast to allow strategic processing, Seymour et al. find nearly identical results even when "guilty" subjects were motivated to appear "innocent" (beat the test).





Method

(Seymour et al., 1997)

1) Memorize the **Probe list**; Study & Recall 3 times.

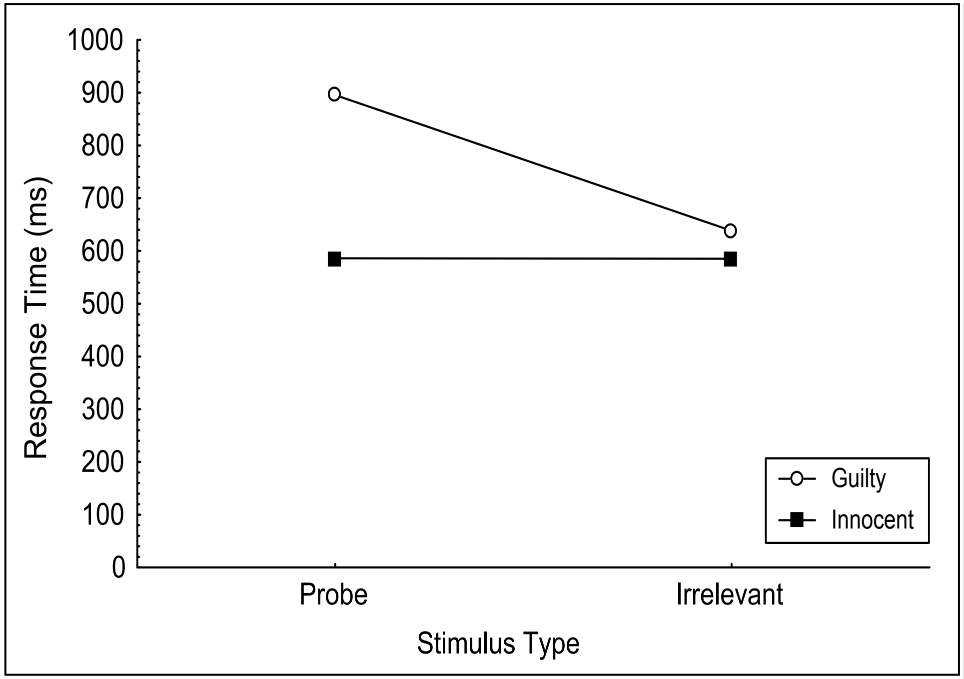
2) Asked (a few refused) to send an email message using **Probe List** by logging into the university computer system as another student and emailing a third party. Email sets up a meeting with the third party. Details of message (e.g., login information & message details) specified by **Probe list**. Subjects are led to believe that this task is a **Crime**, though no message was ever actually sent.

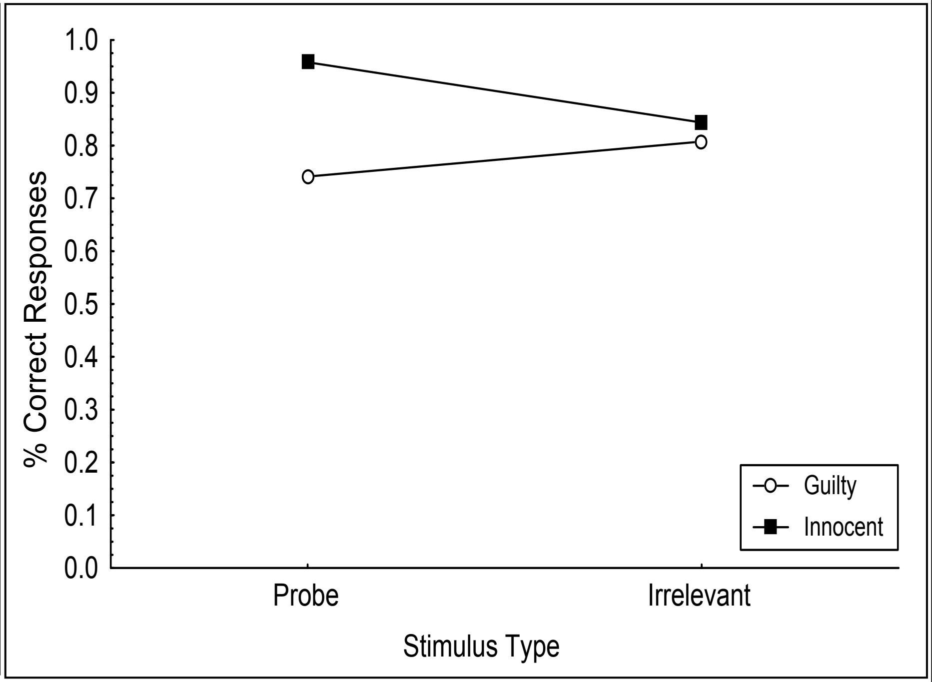
3) **Distractor task** (mathematical word problems) aimed at preventing rehearsal of the **Probe list** and dissociating previous section from the next section.

4) Memorized the **Target list**; Study & Recall 3 times.

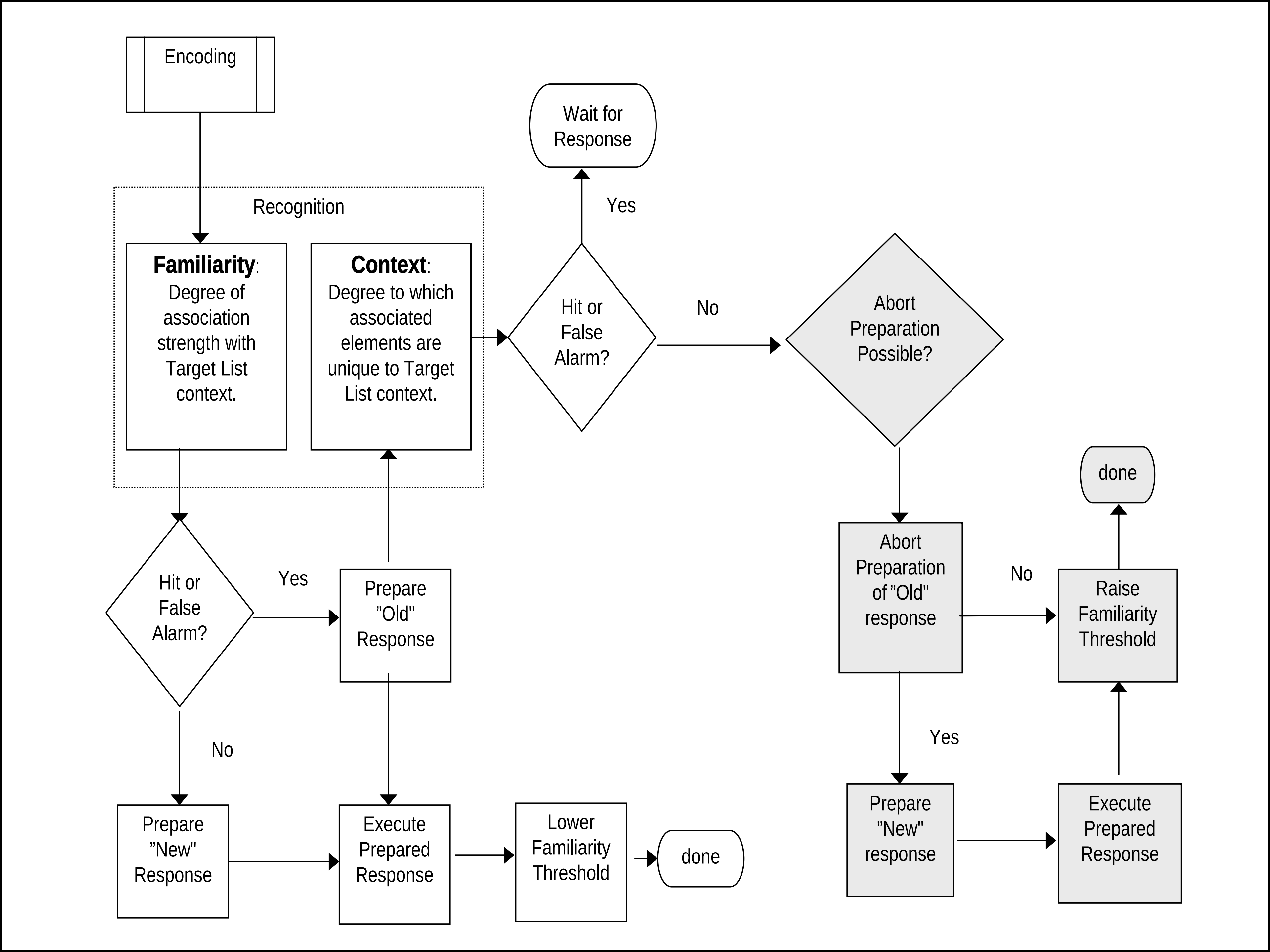
5) Perform **"Old/New" judgment task** indicating whether each stimulus is from the **Target list** or not. Subject unaware that some filler (irrelevant) items are from the **Probe list**. Subjects must reject both **Probe items** and **Irrelevant items** as "New".

Results (Seymour et al., 1997)





"Innocent" subjects who did not commit the "crime" respond identically to **Probe items** and **Irrelevant items**. "Guilty" subjects are considerably slower and less accurate only when rejecting **Probe items**. This pattern is the *Guilty Knowledge Effect*.



Encoding
Recognition
Familiarity: Degree of association strength with Target List context.
Context: Degree to which associated elements are unique to Target List context.
Hit or False Alarm?
Yes
No
Wait for Response
Abort Preparation Possible?
Prepare "Old" Response
Prepare "New" Response
Execute Prepared Response
Lower Familiarity Threshold
done
Abort Preparation of "Old" response
Raise Familiarity Threshold
Prepare "New" response
Execute Prepared Response

A Model of the GKE

Our process model (shown above) incorporates the following components:

1) **<u>Familiarity</u>**: This stage involves a single-process signal-detection type evaluation of the degree to which elements associated with the stimulus are associated with the **<u>Target list</u>** context. This fast and automatic judgment is made without considering the degree to which elements associated with the stimulus are *uniquely* associated with the **<u>Target list</u>** context (e.g., Kintsch 1970; Anderson & Bower 1972). An overt "Old" or "New" response is prepared on the basis of this judgment.

2) **<u>Context</u>**: This stage is characterized by a single-process signal-detection of the degree to which the relationship between elements associated with the stimulus and elements associated to with the **<u>Target list</u>** context are unique to the **<u>Target list</u>** context. This stage occurs concurrently with overt Response Preparation and *only* if the Familiarity stage judges the stimulus to be "Old" (as a check on the accuracy of this judgment). If this stage deems the Familiarity judgment to be sound, the "Old" response is allowed to proceed and the Familiarity threshold is lowered slightly.

3) **<u>Abort Preparation</u>**: If the Context stage reveals that the stimulus previously judged as "familiar" is actually from the **<u>Probe list</u>** and not the **<u>Target list</u>**, the Abort Preparation stage attempts to halt the "Old" response currently being prepared. If this stage fails, an error is made and the Familiarity threshold is raised slightly.

However, if the Abort Preparation stage succeeds in aborting the impending overt response then the appropriate "New" response is prepared and executed.

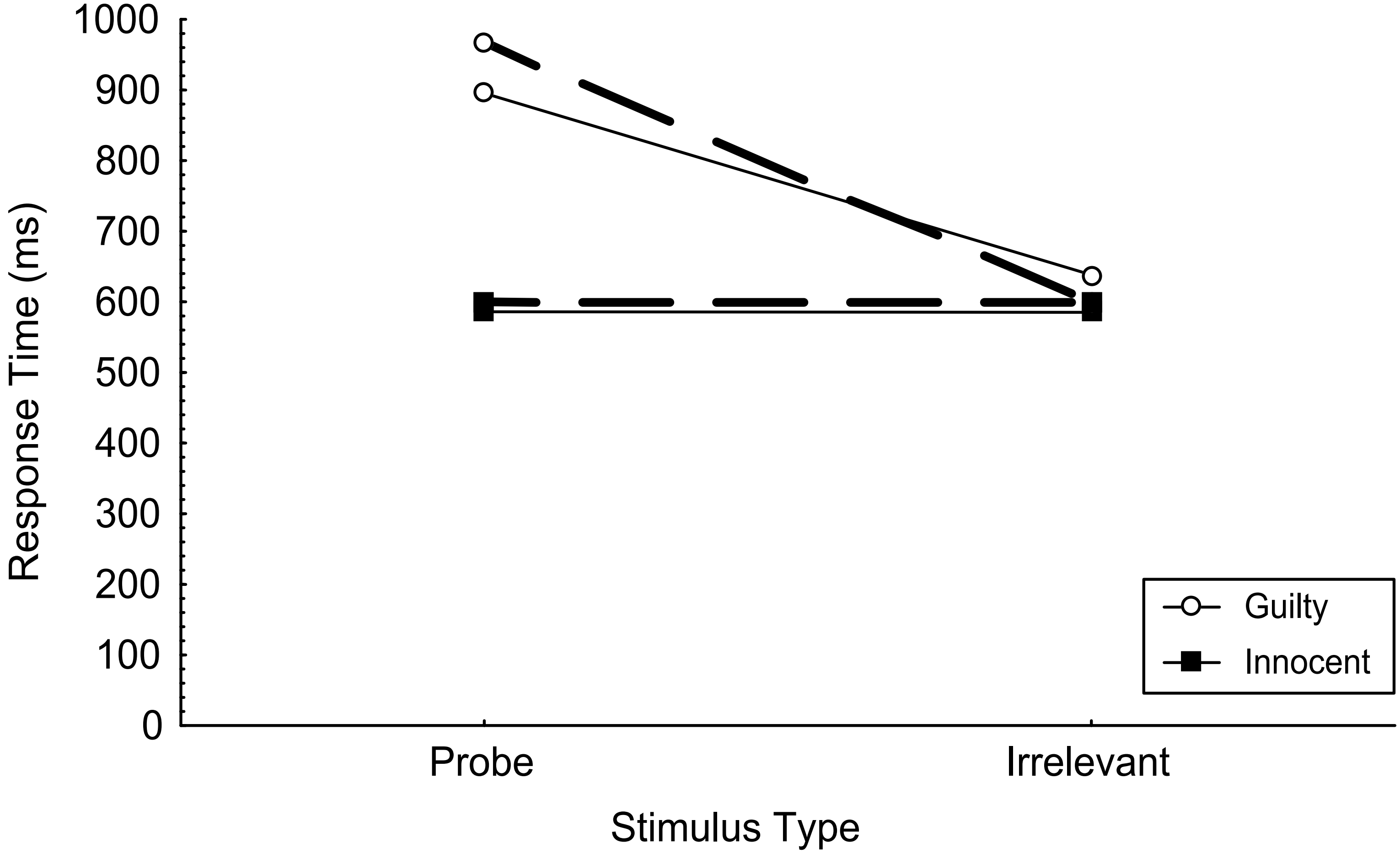
Note that any path through the model that leads to the shaded section will result in an increase in RT. A failed abort will lead to a slightly smaller increase, but will result in an erroneous response. A successful abort will necessitate preparing a new response and will require additional time.

Because only Guilty participants will judge **Probe items** as "familiar" and prepare an "Old" response, they are uniquely doomed to occasionally catch this error which will produce either an erroneous response or an RT considerably slower than responses to **Irrelevant items**.

Key Model Parameters

Stage	**Duration**	**Source**
Encoding + Familiarity	332 ms	Meyer & Schvaneveldt, 1967
Context	218 ms	Reichle, Pollatsek, Fisher & Rayner, 1998
Response Preparation Response Execution Action Initiation Device Transduction Abort Preparation	58 ms 100 ms 50 ms 10 ms 50 ms	EPIC architecture, Meyer & Kieras 1997
Hit or False Alarm Abort Possible	50 ms 50 ms	Based on EPIC architecture, Meyer & Kieras 1997

Predicted Response Time



Note: Dashed lines represent predicted data, Solid lines depict observed data from Seymour et al., 1997.

Modeling Strategic Manipulation

In addition to understanding the processes underlying recognition, modeling the Guilty Knowledge Effect offers an understanding about why certain strategies employed by "guilty" subjects failed and which strategies, if any, would prove more successful.

1) One strategy tried by "guilty" subjects is to set a conservative threshold for the Familiarity stage, reasoning that **Target list** items should have more elements associated with the Target Study context than will any distractor (e.g. **Probe items**). However, because of the high degree of similarity between **Probe** & **Target items**, this strategy is unlikely to be successful, as fine grained contextual distinction will only be available in the later Context stage, concurrent with Response Preparation.

2) Alternatively, subjects may have tried to delay Response Execution until the Context stage had completed, thereby ensuring that aborting previously prepared erroneous responses was always possible. While this eliminates the test's ability to classify subjects on the basis of accuracy, the RT effect remains unaffected.

3) If subjects had postponed Response Preparation until after the Context stage, there would never be a need to take the shaded path, and RT to **Target items** and **Probe items** would be identical. However, because the Context judgment is unnecessary for "New" stimuli, the critical comparison (Probe vs. Irrelevant) would still identify "guilty" subjects.

4) A promising practice regimen would be to contextually elaborate the **Target list** to such an extent that **Probe items** and **Irrelevant items** fail to reach threshold in the Familiarity stage as with "innocent" subjects.

Conclusion

By modeling the Guilty Knowledge Effect we can continue to develop robust tests of Guilty Knowledge, and specify why attempts at strategic manipulation fail. With a detailed model, we can also begin to make predictions about which components of recognition are under strategic control and under what conditions.

Future modeling of recognition may reveal further parameters and constraints of memory retrieval processes, and will help illuminate the limits of conscious control in memory.

References

Anderson, J. R. & Bower, G. H. (1972). Configurational properties in sentence memory. *Journal of Verbal Learning and Verbal Behavior*, *11*, 594-605.

Bashore, T. R., & Rapp, P. E. (1993). Are there alternatives to traditional polygraph procedures? *Psychological Bulletin*, *113(1)*, 3-22.

Farwell, L. A. & Donchin, E. (1991). The Truth will out: Interrogative Polygraphy ("Lie Detection") with event-related brain potentials. *Psychophysiology*, *28(5)*, 531-547.

Furedy, J. J. & Heslegrave, R. J. (1988). Validity of the lie detector: A psychophysiological perspective. *Criminal Justice & Behavior*, *15(2)*, 219-246.

Kintsch, W. (1970*). Learning, memory and conceptual processes*. New York: Wiley.

Lykken, D. T. (1981). *A Tremor In The Blood: Uses and Abuses of the Lie Detector*. New York: McGraw Hill.

Meyer, D. E., & Schvaneveldt, R. W., (1976). Meaning, memory structure, and mental processes. In Cofer, C. (Ed.) *The structure of human memory* (pp. 55-89). S.F. : W. H. Freeman.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, *104*, 3-65.

Reichle, E. Pollatsek, A., Fisher, & D., Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105(1)*, 125-157.

Seymour, T. L., Mosmann, A. L., & Seifert, C. M. (1997). Using Reaction Time to Measure "Guilty Knowledge". *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (p. 1044). Hillsdale, NJ: Lawrence Erlbaum Associates.