ARRESTED DECISIONS: THE EFFECTS OF INFORMATION LATENCY ON HIGH-RISK DECISION-MAKING

By

Kenneth Rosson Pence

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Interdisciplinary Studies: Management of Technology

May, 2005

Nashville, Tennessee

Approved:

Professor David M. Dilts (Supervisor)

Professor Richard Daft

Professor Salvatore T. March

Professor William R. Mahaffey

ACKNOWLEDGEMENTS

This research could not have been completed without the support of my advisor, Dr. David M. Dilts, who provided many insights and recommendations to foster this work. He always encouraged me during the many revisions.

Special thanks to my committee also. Sal March made me think and suggested options that did not hit me until days after meetings in some cases. He suggested that there might be process changes – he was right. Dick Daft was an inspiration to me and saw ideas in minutes that took me days to grasp. Bill Mahaffey encouraged me, taught me systems thinking from a wealth of real-world experience. I would not even be in a position to complete this work without him taking me under his wing.

I would also like to thank my wife, Meredith DeVault, who was sympathetic and made the sacrifices at home to keep me going. An honorary degree should be given to my children for putting up with me in this endeavor. Finally, I would like to thank the Metropolitan Nashville Police Department who provided such a fertile ground for research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

THE EFFECT OF INFORMATION LATENCY ON HIGH-RISK DECISIONS

Introduction

My research evaluates how changes in information latency in computer-based communications affect performance of high-risk decision-makers. Information latency[1], from a decision-maker's perspective, is the time from query to receipt of information. From a baseline of high information latency, the impact of technology insertion with low information latency followed by technology modification resulting in moderate information latency is examined. High-risk decision-makers here are police officers making life and liberty decisions. Two primary theoretical models are used in the research: 1) the recognition-primed decision (RPD) model for high-risk decision-making (Klein 1989)and 2) media richness (Daft, Lengel et al. 1987). Three historical interventions were studied: occurring in 1997, 1998, and 2001. The research design is quasi-experimental, with non-equivalent groups, and non-randomized assignment (Shadish, Cook et al. 2002). Multiple time series observations from approximately 400 subjects have been gathered from multiple archival data sources spanning six years. This exploratory, empirical research is a longitudinal study, using archival data that evaluates the impact of information latency changes on high-risk, recognition-primed decision-making performance. Additionally, my dissertation should refine characteristics within the media richness model and examine decision thresholds.

The following cases give an overview of the problem background. Elements of the recognition-primed decision model, media richness, and information latency are italicized in the following three scenarios, using fictitious names, in police work:

Case One**:**

Officer Johnson, a two year veteran, sees a male subject standing in the dark and lighting a cigarette behind a liquor store an hour after the store is closed. Officer Johnson decides that a man standing alone in the dark behind a business is not typical and may be *analogous* to similar burglaries over the last month so he stops the man. It is Friday midnight. Johnson asks the subject for identification but the man says he has none. The subject has a screwdriver in his pocket. The tool is a *relevant cue* for the officer's theory with recent burglaries as other businesses have had their door locks pried open with a screwdriver. Johnson asks the subject's name and the man says his name is Thomas Smith. Officer Johnson asks the man's date of birth and the subject responds.

The officer switches the radio frequency on his walkie-talkie (hand-held transceiver) and asks to be put in line for records (to be able to ask a civilian employee to query the subject through the use of a computer terminal at a distant location). During the *information latency* delay of nine (9) minutes, the officer waits in a queue with other officers to request information from a records clerk, the officer questions the subject as to why he is carrying a screwdriver in his pocket, where he lives and works. The subject says he is a handyman and had been replacing old light switches at his apartment several blocks away. The ready answer is reasonable and does nothing to clarify the *expectancies* of the officer that the screwdriver was to be used to break into the store. The subject says he is just going for a beer. The records' clerk responds after nine minutes have elapsed from the original stop time. The officer asks for a record check for active warrants[2] and a driver's license check for a valid drivers' license on the subject. Four minutes later the records clerk answers on the radio with a verbal response saying there is no license that corresponds with that birth date for the subject and there are no warrants on file under that name.

The officer asks the man why he has no license and the subject says he never drives a car, which is why he is walking. Again, the officer has no information to confirm the subject's name and it is not unheard of to have no license though this is unusual. The officer seeks *more data* to clarify this *anomaly* so the officer asks the subject for his social security number which the man promptly recites. The officer starts to ask the records clerk by radio for an additional verification check of the subject's social security number but another officer, the next in the radio queue, is reading a seventeen-digit alpha numeric vehicle identification number to the records clerk, to check for a stolen report on a truck in a separate incident. Officer Johnson decides to fill out a field interview report, indicating he has stopped a suspicious person, and lets the man continue on his way, since he has detained him for slightly over fifteen minutes at this point and has no

[1] From a computer standpoint, this is equivalent to twice the network response time plus query processing time.

evidence of active criminal warrants or evidence that the man is committing a crime. Twenty minutes is a generally acceptable length of time a subject can be stopped in the United States without an officer placing charges or have addition evidence to detain a subject (United States v. Sharpe, 1985). The next day, another officer reviewing this field interview report sends Officer Johnson a message saying that the social security number and physical description of "Thomas Smith" matches the social security number and description of a well-known serial burglar who has an un-served, parole violation and burglary warrants on file.

Case Two**:**

A few months later, Officer Johnson decides that a subject's furtive behavior in a shopping center parking lot is *analogous* to behavior he has seen in previous car theft situations so he stops the subject. This man, when asked for identification, claims he has a driver's license but does not have the license with him. Officer Johnson goes back to his car and through his wireless computer terminal, queries this subject's history as a licensed driver and checks for active criminal warrants on the name given. The *information latency* for a driver history response is twelve seconds. This textural response indicates a valid license under that name and a physical description similar to the subject who has just been stopped. Little further information is available and driver's photographs are not available without driving back to the police station. This response does not clarify the officer's *expectancies* that the subject has similar arrests for theft from autos or outstanding warrants so the officer seeks *more data*.

Officer Johnson queries the subject's social security number using his wireless computer terminal and the textural response indicates the social security number belongs to a Hispanic female. This *anomaly* confirms the officer's *expectancies* that the subject is lying and the officer seeks even *more data* to match features which infer that this subject is actively attempting to break into vehicles. The officer further questions the subject and asks when the subject graduated from high school. The subject says, "1989." This response indicates the subject is about twenty-six (26) years old (in 1997). The subject, however, says he is twenty-two years old, the officer perceives this as another *anomaly* inferring deception. The officer questions the subject further. The subject says his friends call him "Bone." The officer is able to query the 'nickname file' on

[2] A criminal warrant is a legal document issued by a competent authority to incarcerate a person, naming a subject with a summary of a crime that this person is purported to have committed (42Am J1st Proc § 2)

his computer. The computer response to the nickname "Bone" displays the name, Adam Scales, and lists an address for a repeat auto thief. The officer asks the subject if he has ever lived at this address. The man confirms that he has lived at that address which aids officers in building the *inferences* that this subject is the named repeat criminal.

The officer asks the subject what he had been arrested[3] for last; a loaded question by the officer inferring more knowledge than the officer has in his possession. The man indicates he has been arrested once for a curfew violation when he was a juvenile. The officer queries this name for active criminal warrants and the computer response displays an outstanding warrant for vehicle theft giving descriptions and addresses that match the subject and where he says he lived. The officer places the man under arrest. The initial behavioral recognition of possible criminal behavior, through careful use of *inference* and additional *clarifying data* led to the arrest. Six minutes have passed since the initial stop of the subject.

The difference in the number of queries between case one (2) and case two (5) may indicate that more queries are made when information latency decreases. The officer in case one had to wait in a queue for minutes to verbally query information he had received from the subject he had stopped. When the verbal response was returned, it elicited a need to verify further information but by this time, another officer's request was in the queue. In case two, Johnson could transmit and receive responses faster; *i.e.* with lower information latency, so he could perform more queries and ask more questions. The quantity of queries in case two indicates an adaptation of the decision-making process to take into account these changes in information latency. An alternate explanation may be that the increased latency in case one, with verbal queries to a remote clerk, causes the officer to reduce the number of questions that could be verified because of the additional time consumed. It is not practical to use a long list of pre-arranged queries for all situations since every situation presents uncertain, dynamic conditions that do not apply to some of the set queries. Time constraints preclude the use of too many queries before time restrictions take effect and prevent further action.

[3] ***Arrests*** are *the taking into custody of persons who have committed a criminal act, a violation of established criminal law.*

Case three**:**

Officer Johnson has been using his wireless laptop to query databases for over three years and his laptop is replaced in 2001 with a newer model. The interface is similar but he notices that the queries take twice as long to return as his older model; i.e. information latency has moderately increased. During his early evening patrol he notices a dark, blue 1995 Saturn sedan being driven slowly by a young, male that looks out of place in this neighborhood where there are often crack cocaine sales. The officer recognizes from *analogous* experience that this person could be attempting to buy crack cocaine and queries the license tag. The local license response takes ten seconds and the officer almost loses the car as other vehicles are entering the street. His previous, older computer would have received a response to this type of query in only three seconds. The license response shows the vehicle belongs to a Johnny Franks in Lavergne, Tennessee, a nearby town. The officer thinks the subject may be attempting to buy drugs because similar occurrences in the same area have occurred during his years of patrol. Johnson queries the name, Johnny Franks, in a local, arrest history database when the car stops at a red light and the response displays six, separate names of "Franks, Johnny" but only one name with a Lavergne address. The response to the arrest query took twenty-five (25) seconds but would have only taken twelve (12) seconds on his previous computer. Officer Johnson lost sight of the vehicle after the traffic light.

In case three the information latency is moderately increased over case two as the latency in case three is ten-fold less than in case one, where there was no computer. What is the quantifiable effect on dozens of high-risk decisions by hundreds of officers during a typical day when the information latency changes? Anecdotal evidence abounds but there is little empirical data describing information latency effects on high-risk decision performance and on the decision process. My research will begin to close this gap in reach.

Research Question and Statement of the Problem

Formally, the question posed is: *How does information latency affect decision-maker performance in repeated, high-risk decisions?*

Understanding the effect of information latency on high-risk decision-making would be very useful in determining if information latency had a *significant effect* on standard performance measures, beyond efficiency measures (Dearing 1990; Jamieson 1996; Kalakota and Whinston 1996). For example, designers of decision support systems (DSS) in emergency management situations where time-pressure and high-risk are involved (Banks, J et al. 1996; De Silva 2000; Srinivasan, Sundaram et al. 2000) would benefit from research demonstrating the effectiveness of different levels of information latency.

It would also be useful if it were shown that information latency had *no significant effect* on high-risk decision-making as this would cast doubt on the importance of rapid feedback in many communication and decision models. It would further decision-making research by showing how high-risk decision-makers adapt their cognitive processes according to the information latency when they make their decisions. Such adaptation has been shown with time pressure alone (Johnson, Payne et al. 1993) but my research demonstrates query adaptation, by multiple cues, to a combination of time pressure and high-risk. A model from my research on information latency effects may provide explanations of decision processes in other situations where high-risk decisions are prevalent.

One other interesting area of investigation in this dissertation is an investigation of possible non-linear impact of information latency changes through their impact on performance based on the ability to perform multiple, low latency queries. By non-linear impact, I mean that there may be a range of query types and quantities of queries that demonstrate a significant change in performance measures after a certain threshold number of queries or query types. It may be that there is a certain reduction in the quantity of queries that causes a drop in decision-making performance. This is referred to as a threshold effect.

This threshold effect may vary depending on the reductions in latency under certain conditions. A reduction in latency may be statistically significant in one situation but may have a statistically significant effect on performance in one situation but not in another. This study will also examine if conditions affect thresholds.

Definitions

Latency terms defined:

- ***Latency*** is the "time it takes a signal to propagate from the sending node to the receiving node once it has been transmitted without any consideration of processing." (Johansson, March et al. 2003)
- ***System Latency*** is the sum of all the network communication delays
- ***Network response time*** is the combination of: transmit time, queuing and intermediate processing delays plus system latency. (Johansson, March et al. 2003)
- ***Information latency***, from the decision-making perspective, is the time from query to receipt of information. From a computer standpoint, this is equivalent to twice the network response time (query request message and query response message) plus the time to process the query.

Query terms defined:

- ***Quantity of self-initiated queries returned*** is the actual number of queries the user in the field has sent and received in a defined time period.
- ***Warrant check*** is a query looking for un-served criminal warrants for a particular person through local, state and federal databases.
- ***Arrest history*** is a query that shows if a particular person has been arrested. This query also shows the criminal charges and disposition of the arrest.
- ***Drivers' license check*** refers to a query that will return if a particular person has a valid issued driving license including any restrictions, revocations and history of citations or accidents.
- ***License check*** refers to a query checking local, state and federal databases by the numbers on a vehicle's license plate to see who has registered the vehicle and if it is reported stolen. Stolen checks are the same with any manufactured goods with a serial number or owner applied number (there are many types of stolen queries).
- ***Nickname file check*** refers to a query looking for alternate names used by a particular subject or names this person is called by others and cross-referenced to their legal name.

Naturalistic decision-making definitions for my research:

- ***Naturalistic decision-making* (NDM)** is the way people use their experience to make decisions in field settings (NDM conference 1994). NDM is defined by eight characteristics: ill-structured problems, uncertain dynamic environments, shifting and competing goals, action-feedback loops, time stress, high stakes, multiple players, and organizational goals and norms (Orasanu and Connolly 1993; Drillings and Serfaty 1997)
- ***High-risk decisions*** are defined as those decisions where there is a possibility of loss of life or serious bodily harm, loss of freedom (incarceration) or freedom, forced compliance or negotiation, or other serious consequences due to the decision.
- ***Experienced decision-makers*** are decision-makers who have encountered significant levels of analogous situations within a field over a period of time, usually measured in years.
- ***Time pressure*** means having a short period of time, either seconds or minutes, to reach a decision and in this dissertation is defined by a legal threshold of articulatable information where a reasonable person would believe conditions had been met or from a deadline imposed by legal conventions (approximately twenty minutes in police work).

Decision-making model attributes:

- ***Analogous*** situations are those situations that are similar to past experiences.
- ***Prototype*** situations are novel or new situations without similarities to previous situations.
- ***Expectancies*** are circumstances or features of a situation that are anticipated.
- ***Cues*** are indications or prompts in a situation.
- *An* ***anomaly*** is an "expectancy" that is missing. It is also when unexpected cues are present or circumstances are different than expected.
- ***Diagnosis (or story building)*** is the *clarification of anomalies* by inferences or gathering *more data*. This also refers to feature matching with a mental 'picture' of what is happening.
- ***Situational awareness*** is the internal *story building* by an experienced decision-maker experiencing a situation in a changing context (Endsley 1995). Another way of describing

this term is a mental "template" of cues being checked by the decision-maker in a dynamically changing situation.

- ***More data*** means gathering additional information about a situation to clarify a mental story.

Other decision terms:

- ***Equivocality*** means ambiguity, the existence of multiple and (potentially) conflicting interpretations about an organizational situation (Daft and Lengel 1986).
- ***Uncertainty*** refers to missing information, conflicting information or complex information (Daft, Sorumen et al. 1988; Schmitt and Klein 1996; Cole, Vaught et al. 1998).

Threshold decision terms for police work:

- ***Reasonable suspicion*** is the legal threshold of information needed by an officer to stop a person for further questioning. Reasonable suspicion to stop a person for an investigatory interview exists when an objectively, reasonable officer perceives events that would cause them to suspect criminal activity on the part of the individual stopped (App. 2001).
- ***Probable cause*** is the legal threshold of information, more than reasonable suspicion, needed by an officer to arrest or place a person in custody. Probable cause to arrest exists when an objectively, reasonable officer has articulatable facts and circumstances to believe that a crime has been committed and this person is believed to be the person that committed the crime (Brinegar v U.S.,1949; Draper v United States, 1959).

Generic threshold decision term:

- D***ecision thresholds*** are the levels of accumulated information necessary to reach a decision point.

CHAPTER II

LITERATURE REVIEW

People make decisions in non-staged, real-world settings daily. Most of these decisions are not carefully weighted but use a singular evaluation strategy (Simon 1957). Experienced decision-makers often arrive at a solution using a pattern-recognition process that is non-linear (Tversky and Kahneman 1974; Klein 1989; Swets 1992; Klein 1998). The pattern recognition process involved has been modeled in naturalistic decision making by including high risk and time pressure as components that are not included in classical, rational choice decision theory. Literature on high-risk decision-making differs greatly from decision-making with little risk, hence this chapter focuses on high-risk decision-making (Klein, Orasanu et al. 1993).

High-Risk Decision Making

High-risk decisions are defined as those decisions where there is a possibility of loss of life or serious bodily harm, loss of freedom (incarceration), forced compliance or negotiation, or other serious consequences due to the decision (often made under time-pressure). I examine *high risk* decisions, not high stakes. *High stakes* have been described as the stakes in gambling using large and small wagers with little consequential risk[4] (Ordonez and Benson 1997). Decisions examined in my research are more similar to high-risk decisions made at a nuclear plant (Miao, Zacharias et al. 1997). The consequences are lives at risk. Classical, rational decision models do not accurately portray real world decisions with high risk (Allison 1971; Allison and Zelikow 1999) where naturalistic decision making models often include severe risk (see below).

High-risk decision-making research has been studied in firefighting (Weick 1993), a nuclear plant during emergency operation (Miao, Zacharias et al. 1997; Hoffman, Crandall et al. 1998; De Silva 2000), military operations in missile defense (Lanir 1989; Kaempf, Klein et al. 1996), pilots with uncertain information (Fischer, Orasanu et al. 1995), tactical operation commanders (Schmitt and Klein 1996), doctors in an emergency room (Crandall and Getchell-Reiter 1993), and other decision-makers under time pressure in high-risk situations (Banks, J et

[4] The risks of wagering money are not life-threatening even though there are significant decision process differences between large and small wagering bets.

al. 1996; Bechara, Damasio et al. 1997; Klein 1998; Maule, Hockey et al. 2000; Srinivasan, Sundaram et al. 2000). Much of the research data from actual crisis situations, however, is unreliable or fragmentary due to the small sample sizes[5], chaotic communication during incidents and poor identification of factors affecting the incidents (Lanir 1989; Ross and Staw 1993; Fischer, Orasanu et al. 1995; Miao, Zacharias et al. 1997; Schaafstal, Johnston et al. 2001; Roper 2002). The primary research methods used in these past studies are based on experienced persons building a 'story' from analogous previous situations and then testing those analogies. Typically, studies of crisis or critical decisions are conducted through interviews with participants or witnesses.

There are always serious issues with using post hoc verbal reports as data (Ericsson and Simon 1993). Ericsson and Simon (1993) posit that verbalization is an imprecise method to transform behavior into data since encoding verbal protocols as data suffers from interpretation of verbal response. Persons have difficulty remembering multiple, spoken details unless these details are saved in written or electronic text. They further posit that this interpretation of behavior may also suffer from introspection. Research using archival, empirical data, as I used in this dissertation, is much less likely to suffer from interpretation errors as when attempting to turn verbal data into 'hard' data. The experience of the decision-maker appears to affect decisions differently from casual participants in high-risk decisions also.

Experienced Decision Makers

Experienced decision-makers choose alternatives differently than novices. De Groot (1986) researched chess players of different skill levels and found that the more expert players did not look at all the alternative strategies before a move but instead their first choice was significantly the better choice. The more experienced the player, the more likely that player was to pick a move and assess whether that was the move to make. It was the novices that would try to weigh a large number of alternatives. Both novices and experienced persons in novel situations proportionally make worse decisions when they perform rational choice strategies first (multi-attribute utility analyses) and try to consider all alternatives (Erev, Bornstein et al. 1993). This would mean that if a decision-maker had a series of analogous experiences they would be able to compare (recognize) similarities from their earlier experience to what is happening

[5]Often a low number of persons interviewed after incidents

without trying to weigh so many alternatives. It would also mean that novices and experts in new situations might be on a similar footing in information gathering though experts have been shown to gather more information than novices in situations outside their experience (Kardash, Royere et al. 1988). Decision-makers under time-pressure appear to use different decision-making strategies.

Time Pressure

Time pressure, when there are 'lives on the line', is not found in classical decision theory and is even ill-suited for routine naturalistic decision-making. For example; Svenson and Benson (1993 in Maule) cite that decision effectiveness in making choices, decreases under time deadlines in non-crisis events. Decision quality has also been shown to be reduced under time pressure (Johnson, Payne et al. 1993; Svenson and Benson 1993). Johnson, Payne and Bettman (1993) showed that decision-makers adapt to time-pressure constraints using three general adaptation strategies. They posit decision-makers 1) try to accelerate decisions using the same processes, 2) become more selective in information used to make the decision when time pressure increases, using subsets of the information they would normally use; or 3) adapt to a different decision-making strategies (equal weight, framing, satisficing, etc.) than the decision strategy they were currently using and they make better decisions than those who don't adapt strategies (Johnson, Payne et al. 1993). I examine instances of serious time pressure with possible deadly consequences. Time pressure has been shown to distract, interfere (cause noise), and limit information gathering (Wright 1974; Pennington and Hastie 1993).

Naturalistic decision models are a type of model that encompasses the elements of high-risk, time pressure and experience. Naturalistic decision-making (NDM) is designed for real-world decisions and is not linear decision-making. Classical, linear decision-making does not lend itself well to the real world as situations are dynamic and information is incomplete or uncertain (Cohen 1993). Classical, rational decision making is more a comparison of the possible choices at one time while naturalistic decision-making examines situations by identifying an appropriate choice from their experience and then assessing that choice using feedback (Beach and Lipshitz 1993) hence I must adopt a decision model like NDM rather than a classical model. The experienced decision-makers examined, police officers, have ample analogous experience to compare to situations they encounter.

Selection of the Recognition Primed Decision Model

People, in non-staged settings, rarely make deliberate, reasoned choices of many alternatives but instead make decisions based on their experience and intuition (Crandall and Getchell-Reiter 1993; Bechara, Damasio et al. 1997). The recognition primed decision model (Figure 1: Inside the RPD model) is a naturalistic decision model that matches the characteristics of the participants in this study: high-risk, time pressure, and experienced decision-makers in a dynamic environment. Recognition-primed decisions are made by experienced individuals, not novices in unfamiliar settings. Research has shown that the experience of the decision-maker affects decision quality (Kahai and Cooper 2003). RPD suggests that persons recognize an *analogous situation* from their experience using *cues* from the environment (setting) and matches these cues to a plausible story of what is happening (Orasanu and Connolly 1993; Weick 1993; Schmitt and Klein 1996).

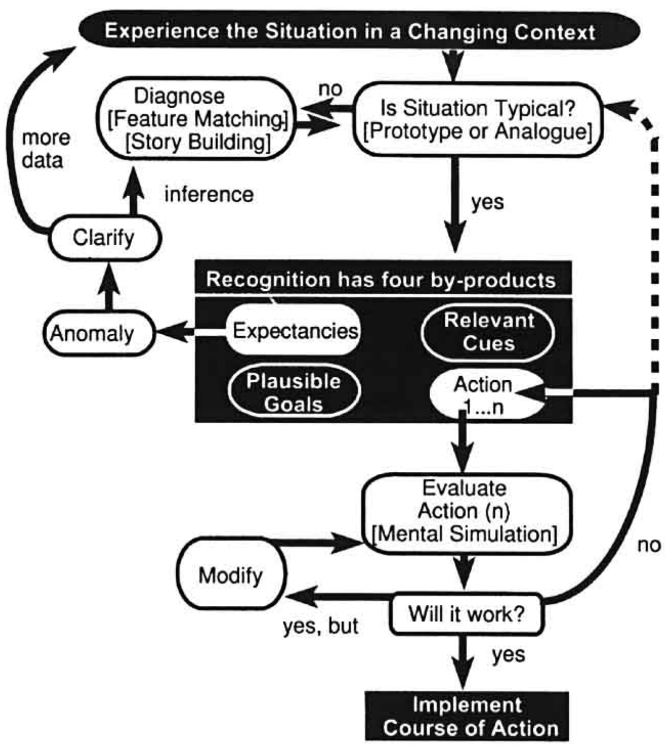

Figure 1: Inside the Recognition-Primed Decision Model (Klein 1998)

The three cases in the introduction of this research describe how individual officers recognize possible criminal behavior that fits a previous pattern from their experience or training in the seconds before stopping that individual. They cannot stop a person without being able to articulate the reason. The officers then look for cues, plausible *goals* and *expectancies* that match their original recognition of the situation. Such *cues* and *anticipated actions* in the RPD model are influenced by additional information that either reinforces or modifies the *story* for the decision-makers. In the example cases, uncertainty appears to be a barrier to action (Daft and Macintosh 1981; Orasanu and Connolly 1993) and *more data* is sought in the recognition-primed decision model to reduce uncertainty, missing information, or equivocality, to clarify conflicting information, and increase decision quality.

The recognition-primed decision (RPD) model is well tested as a NDM model in high-risk situations; it combines recognition, story building, and more data for clarification into a unified model (Pascual and Henderson 1997). The recognition primed decision model can be termed a three 'variant' model. Each variant is a different path through the model as illustrated in figures 2 through 4 and described below. In the first variant (Figure 2), the decision-makers in the RPD model recognize familiarities in a situation from their experience.

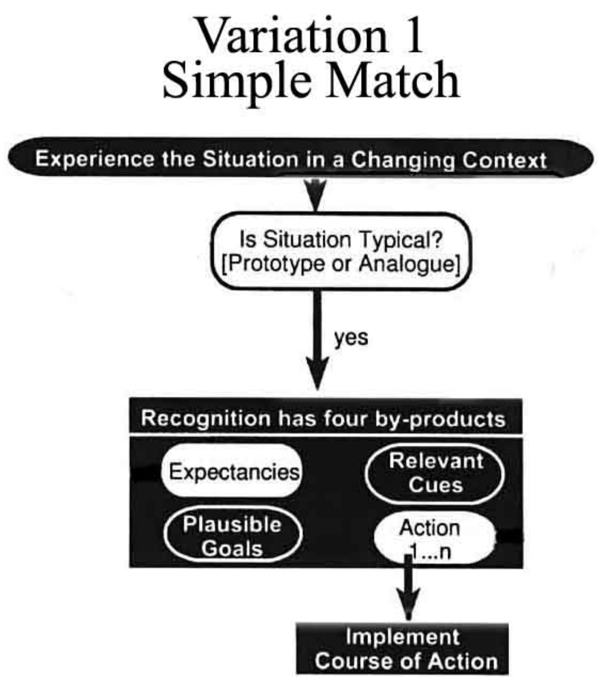

Figure 2: RPD Simple Match Variant (Klein 1998)

They then implement a course of action. This variant does not require additional data for clarification (see Figure 2).

In the second variant (Figure 3), they may recognize similarities from several previous situations or have a complex situation that does not quite fit their previous experience and have to spend more time diagnosing that situation. If expectancies are not met they will try to resolve any anomalies with more data. This is what officers did in making the decision to arrest in the first two example cases. Only variant two (Figure 3) requires additional data for clarification for the decision-makers.

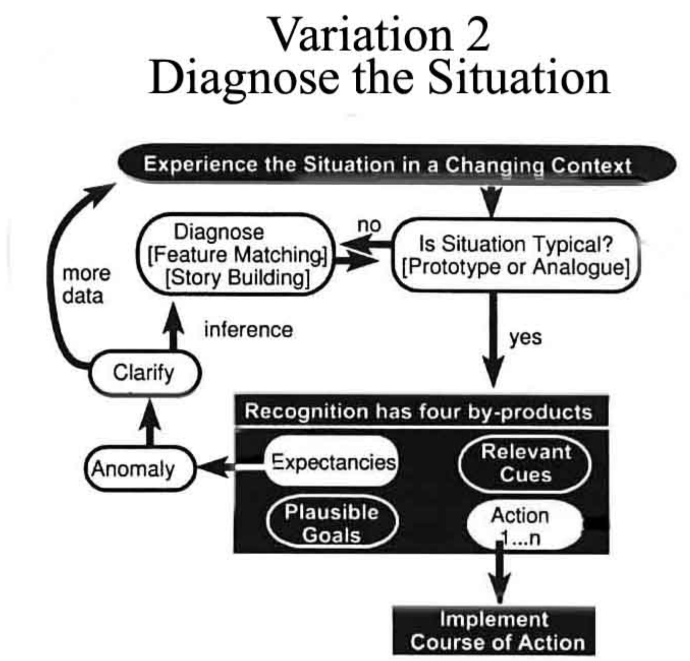

Figure 3: RPD Diagnostic Variant (Klein 1998)

The third variant models how decision-makers may mentally anticipate difficulties with the implementation of their actions and simulate workable actions. This variant does not require additional data (see Figure 4).

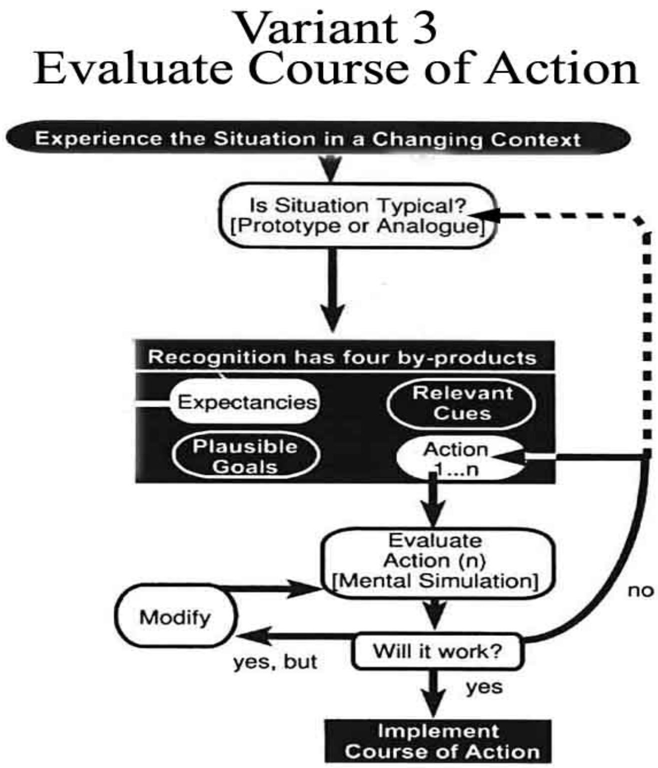

Figure 4: RPD Mental Simulation Variant (Klein 1998)

<u>More data are necessary</u>

The officers in the example cases clarified expectancies and cues with additional queries or interview questions provided by their recognition of typical situations. This clarification is generally necessary after the stop because the justification to arrest must rise to the probable cause standard[6] (Criss v. City of Kent 1988; 2000). The decision to arrest may arise rapidly but must always reach the status where a reasonable officer believes that the person stopped is committing or has committed a crime or is named in an outstanding warrant (legal document). This arrest decision can be decomposed using the RPD model. In its simplest form, the RPD model is a two-stage model with recognition preceding an action but complexities, inconsistencies, and uncertainties may invoke diagnosis that requires more data and more cognitive processing after the recognition phase.

[6] Where a reasonable officer would believe that a person had committed a criminal violation of law

A deficiency in the RPD model is the lack of a provision to demonstrate evidence of decision performance, though it models actions taken after a decision. These *actions* have measurable outcomes. If *more data* are needed to clarify its feature mapping, or situational analysis (story building); then any change in information latency should affect decision outcomes. High information latency may result in fewer queries and failure to reach the level of clarification necessary to rise above the *reasonable suspicion* standard to reach the decision threshold necessary to arrest – *probable cause*. Failure to clarify anomalies in recognition due to information latency (reduced numbers of queries at the aggregate level for many decisions) should then reduce actions. Increased clarification through reductions in information latency should increase the number of queries that produce actions.

The key for my research is the diagnostic process of recognition-primed decision (RPD) model dealing with *more data* and is examined by testing for relationships between changes in aggregate numbers of queries and performance outcomes. Low information latency (high speed query response relationships) may increase the number of queries. The diagnosis portion, in the upper left of the RPD model, is tested where changes in information latency should affect changes in the number of queries necessary for clarification causing resulting changes in decision outcome measures.

Proposition 1: *Information latency has a significant impact on decision performance during high-risk decision-making.*

This would tend to enhance the importance of information latency whenever high-risk decisions are made and allow a direct quantification of information latency to performance measures.

Impact of Media Richness

The media richness model is also examined because it suggests that media that is "richer" is better able to reduce equivocality or ambiguity (Daft and Lengel 1986; Daft, Lengel et al. 1987; Daft, Sorumen et al. 1988; Rice 1993). Media richness theory posits that a combination of four criteria of information in any media: feedback, multiple cues, language variety, and personal focus are used to reduce equivocality (Figure 5). Feedback, as described by Daft, would be affected by changes in information latency. Daft (et al) spoke of 'instant feedback' when referring to interactions with other persons and this reference to 'instant' may be the key to feedback having an effect. The media richness model has been extensively tested with but there has been little empirical support for the model affecting performance (Fulk and Boyd 1991; Carlson and Zmud 1999). Dennis and Kinney (1998) examined the reason for this lack of support suggesting that when examining media richness theory, researchers examined media choice rather than media use (Dennis and Kinney 1998; Dennis and Valacich 1999; Higa, Sheng et al. 2000). More recent research has shown that two characteristics of media richness; feedback and multiple cues, demonstrate positive performance impact with richer media (Kahai and Cooper 2003). One portion of Kahai and Cooper's research even used multiplicity of cues and immediate feedback to determine if these characteristics were better to detect deception and expertise between study participants and showed modest performance gains but their sample sizes are small. Testing for performance changes with these two media richness characteristics with larger sample sizes in another context may provide more insight.

High-risk decisions are by nature, critical decisions where uncertainty tends to block action (Lipshitz and Shaul 1997). Media richness theory research suggests that the multiplicity of cues and feedback (Daft and Lengel 1986; Kahai and Cooper 2003) factors may explain how decision-makers test the *story* they have been building in their mind. The RPD model is unclear as to what constitutes "more data." It may be that media richness attributes may be used to describe how information attributes improve decision quality by reducing equivocality. I intend to test the interaction of these theories to see if it is the combination of the media richness criteria or a single criterion (feedback defined by the number of queries) that demonstrates the greatest, or any, effect on high-risk decision performance. The *multiple cues* characteristic of media richness is tested by examining the effect of queries of different types (arrest history, driver's license information, outstanding warrants, nickname search, juvenile history, and others) on arrest performance. By this I mean if an officer in low information latency settings (fast query responses) changes the way they clarify anomalies with queries of ten or more different types rather than using three or four types of queries. The media richness model uses different types of media while my study only compares voice or self-initiated textural queries. In my study the types of textural queries varies from four generic, multi-part queries to additional queries that provide additional, specific information about a subject, vehicle or situation.

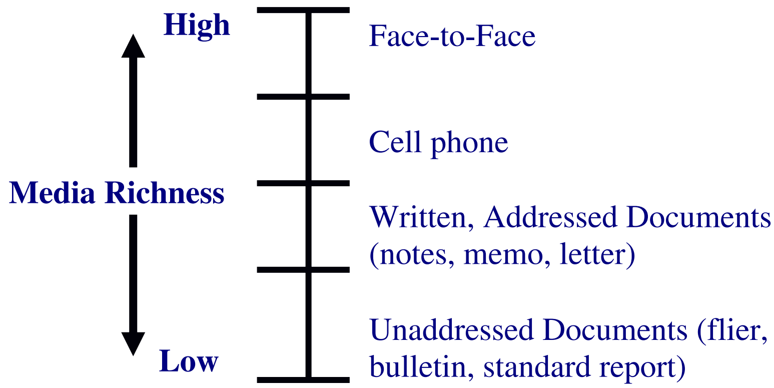

Figure 5: Media Richness - (Daft, Lengel and Trevino 1987)

The RPD model uses *more data* to clarify equivocality and the media richness model uses "instant feedback so questions may be asked and corrections made (Daft, Lengel et al. 1987)." Kahai and Cooper (2003) chose feedback and multiple cues because they considered these characteristics the most important to richer media (Kahai and Cooper 2003). This leads to the proposition that media 'richness' characteristics of feedback and multiple cues should be examined in their relationship to high-risk decision performance, particularly changes in the number of queries due to possible information latency changes. The RPD model provides a model wherein decision-makers build a near immediate analogy from their experience (story building) and then test that *story* with more data. They have expectancies that they anticipate and reevaluate when those expectancies are missing (Shanteau 1992). The media richness portion characteristics focuses on the feedback (number of queries) or the 'multiple cues' (multiple types of queries) characteristic of *more data*.

Proposition 2: *Media of different 'richness' has a significant impact on decision performance outcome measures during high-risk decision-making.*

Such a finding would focus the importance of media richness; particularly feedback and multiplicity of cues whenever high-risk decisions are made. There may be a cumulative effect when data with different richness characteristics or a significant change in performance measures is demonstrated after a certain threshold of queries (see figure 6).

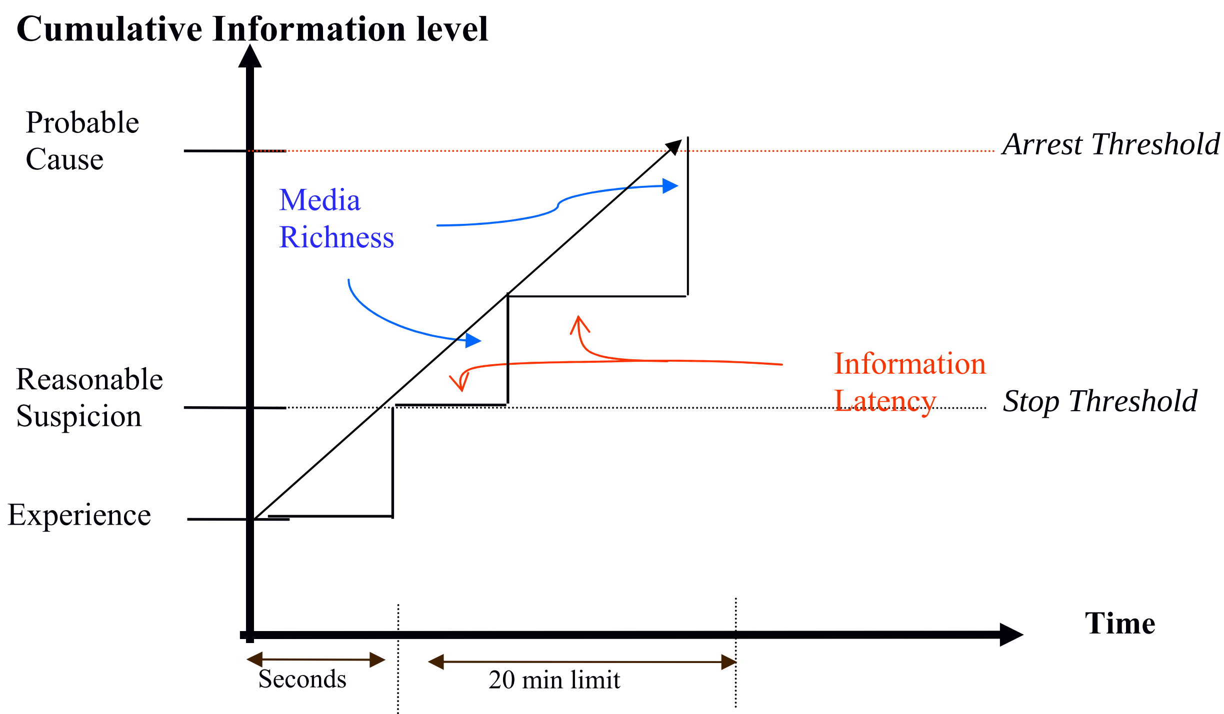

Figure 6: Individual arrest decision-making scenario

Threshold Impact of Information Latency

Research has proposed a decision threshold with information to reach a level of reasonable suspicion (Weaver and Richardson 2002). Decision thresholds have been shown in negative decisions in judicial evaluation (Gilliland, Benson et al. 1998) and in high-stakes diagnostics between binary choices (Swets 1992). Keller and Staelin (1987) demonstrated that more information and information of different types improves decision quality, up to a point, in their study of consumer choices. There may then be a threshold effect, a non-linear relationship, between the number of queries and decision outcome measures. Non-staged decision-making is non-linear since decision-makers do not carefully weigh alternatives and often arrive at a solution using a pattern-recognition process using non-linear "jumps" to first solution that seems to satisfy (Simon 1957; Tversky and Kahneman 1974). Legal standards impose decision thresholds for actions with the decision-makers chosen in this research; police officers. There may be a threshold relationship with number and types of queries due to changes in information latency. The decision makers must reach a threshold of *reasonable suspicion* to stop a person. Then they must reach a higher threshold of *probable cause* before they may arrest a person.

Proposition 3: *There is a significant impact on decision performance outcome measures during high-risk decision-making when several types of queries are "layered" or presented together (possible cumulative, non-linear effect).*

This would tend to enhance the importance of melding data of different types whenever high-risk decisions are made. The model is summarized in figure 7.

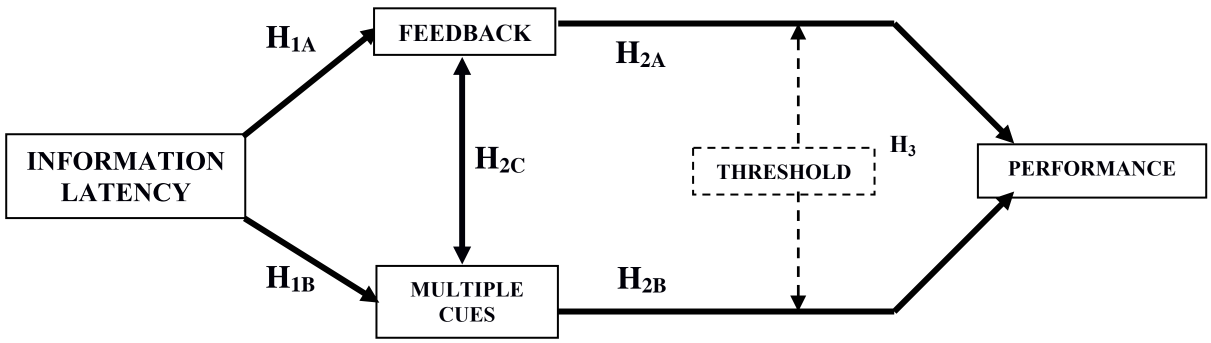

Figure 7: The Research Model

H_{1A} - *Information latency affects feedback in high-risk decision-making*.
Faster responses lead to more queries (mean queries per day).
H_{1B} - *Information latency affects multiple cues in high-risk decision-making*.
Faster responses lead to more types of queries.

H_{2A} – *Feedback, by the number of queries, affects high-risk decision performance*. H_{2B} – *Multiple cues, by the number of types of queries, affect high-risk decision performance*.
H_{2C} – *There is an interactive effect between feedback and multiple cues in high-risk decision-making*.
The number of queries (mean queries per day) and types of queries from different databases allows more clarification of anomalies in high-risk decision-making leading to better performance. These hypotheses test the media richness component effect on performance as well as clarification of anomalies with the recognition-primed decision model.

H_3 - *Information thresholds moderate performance in high-risk decision-making*. The relationship between the intermediate (mediating) variables and performance is non-linear.

The following figure (figure 8) graphically demonstrates three separate cases of using cumulative information to make an arrest. Information latency is portrayed on the X-axis for the different queries. The horizontal component of the queries may be the media richness characteristics of multiple cues and feedback, operationalized as queries to different types of textural information and the mean numbers of queries per day in this research. Cumulative information thresholds that do not reach a level of probable cause do not lead to an arrest. The information latency of multiple queries may preclude reaching an information threshold of probable cause. If information latency is low (fast responses to queries) then the threshold of information, probable cause, should be easier to reach when there are time constraints. Conversely, the threshold of probable cause should be more difficult to reach if information latency is increased (slower responses to queries).

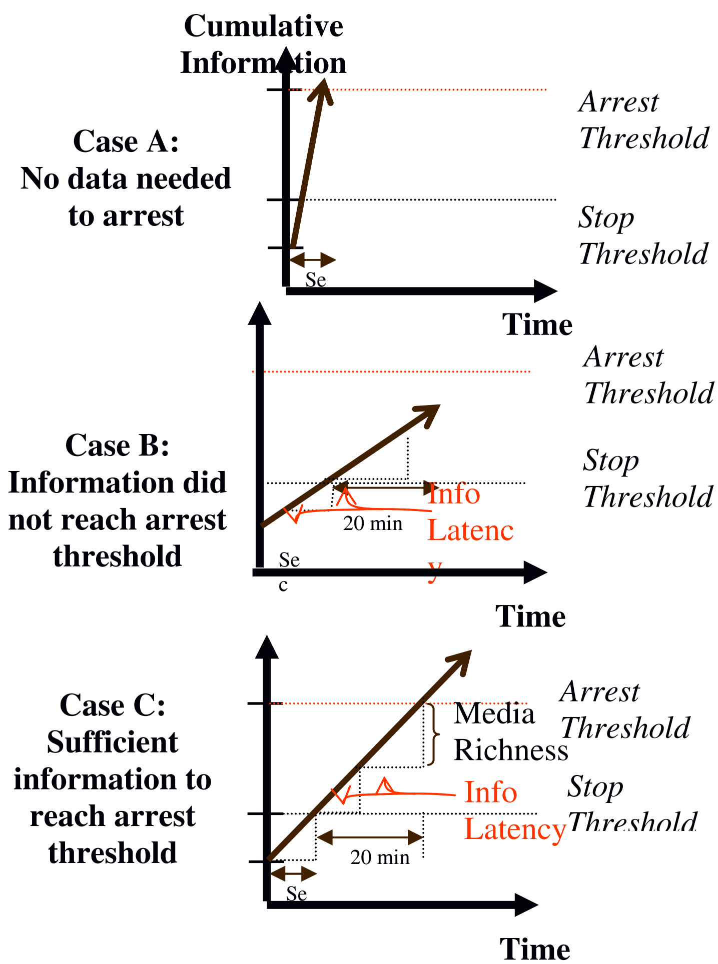

Figure 8: Three example cases of cumulative information gathering

Case “A” corresponds to a situation in which the officer recognizes a subject as a wanted subject, without additional data, or has sufficient facts or circumstances to believe a subject has committed a crime without needing additional clarifying data.

Case “B” corresponds to a situation in which the officer queries for information to resolve anomalies but either did not collect enough clarifying data or exceeded the time limit to gather more data.

Case “C” corresponds to a situation in which the officer queries for information to resolve anomalies and receives enough information to reach an information threshold of probable cause to arrest.

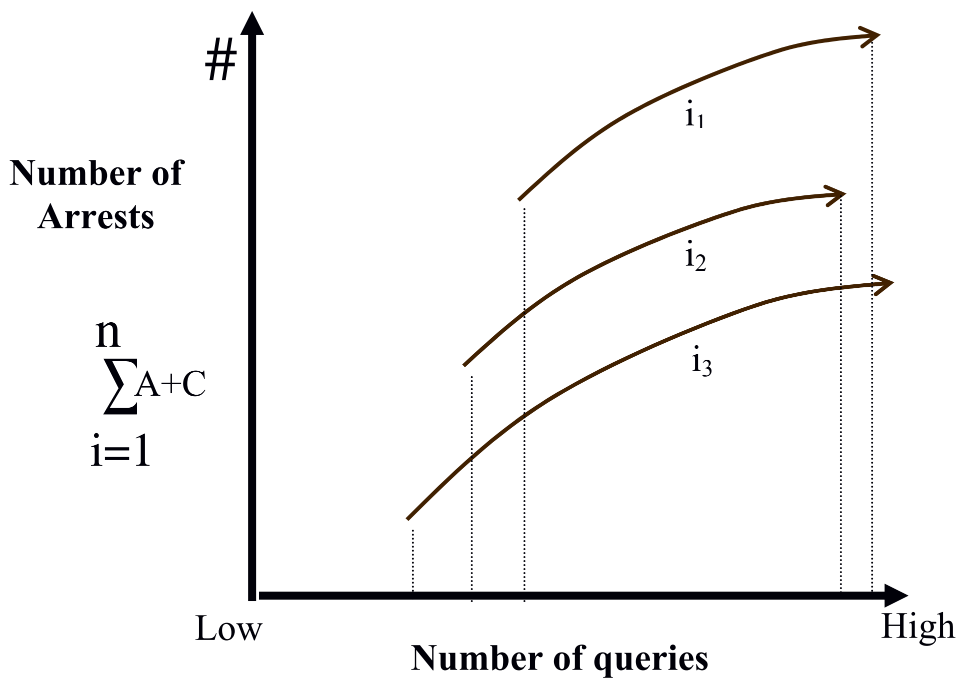

Figure 9: Potential Relationship between Aggregate Decisions and # of Queries

The aggregate case decisions of an officer might be summarized by the graph in figure 9 above. This aggregate includes case A, B, and C decisions and more information was not needed in Case A as information was already sufficient for arrest. Case B there was no arrest and Case C the addition of information resulted in more arrests. Changes in information latency should allow changes in the mean number of queries per day. More queries should result in additional arrests.

CHAPTER III

RESEARCH DESIGN

Decision-Making Setting

The research setting uses street police officers making decisions in a high-risk environment whether to stop and arrest individuals they encounter who may be or not be the subject, listed person, for a criminal warrant[7]. The officer must make the initial decision to stop a person under time pressure due to legal requirements (United States v. Sharpe, 1985) and the decision is normally based on incomplete and often contradictory (equivocal) information. The officer must make a decision using her/his experience, intuition and the additional information gathered at the site as to whether this person they encounter is to be stopped and/or arrested (Criss v. City of Kent 1988; Gardenhire v. Shubert 2000). The consequences of an incorrect decision are severe with civil and criminal penalties in some cases (civil penalties or imprisonment for the decision-maker). In Tennessee, the location of this study, the person being arrested may resist arrest to the point of using deadly force if they consider the officer's force used to arrest them is excessive (TCA 39-11-611 (e) 2). These decisions are not solitary, single examples of judgments but are decisions repeated several times a day – every work day, with only seconds to reach a decision threshold of reasonable suspicion and twenty minutes to then reach a decision threshold of probable cause (United States v. Sharpe 1985).

The police officers studied are faced with dynamic situations and must be aware of possible threats and actions occurring around them. Officers may not initiate a stop without having articulable reasons to suspect that a subject has committed or is committing a criminal violation. This reasonable suspicion occurs in the seconds of observation of an individual or an individual's behavior prior to the stop or detention of that subject. The individuals these officers encounter, in adversarial interviews, are not inclined to readily respond to questions, so officers must use outside sources of information to corroborate subject responses with query responses from outside databases; *i.e.* feedback. Officers are trained to examine a subject's behavioral pattern and relate the subject's current behavior or responses to similar situations from the

---

[7] A warrant is a legal document issued by a competent authority to incarcerate a person, naming a subject and summary of a crime that this person is purported to have committed (42Am J1st Proc § 2).

officers' experience. The officer then asks questions that confirm or refute the subject's guilt. Officers preparing to stop a subject must make an initial assessment, based on their experience before the stop, and then gain information during a stop or interview that confirms or contradicts their earlier assessment.

Officers in this study are experienced decision-makers because, by department policy, they are not allowed to work alone until trained and experienced for at least eleven months. Experience requirements are a well-defined standard across police departments in cities with a population over 100,000 residents in the United States. Three historical interventions that separate the changes in information latency, in a real-world setting, are used.

The officers' decision processes and support system treatment are examined comparing three methods of information retrieval: 1) a radio to query a third party with access to a computer terminal who would then verbally read the textural data returned (high latency) or 2) through use of a wireless computer system allowing the officer to perform self-initiated queries and display the same type of textural information (low latency) and 3) by changes in information latency of the computer system through different communication technologies (moderate latency).

The three types of information retrieval, or methods of gathering more data, are described in the following. 1) The verbal feedback from queries is simplified in this study by examining how remote operators read query responses back to users over a radio transceiver system that does not allow for interruptions. Further the voice responses are limited by reading the textural responses on a remote terminal similar to automated appointment reminders from forced scripts that have limited 'richness' from multiplicity of cues (tone, inflection, etc.). Feedback, in the form of an abbreviated script, is also very limited due to time pressure resulting from multiple queued users striving to obtain their own information through a single source. 2) The self-initiated queries are aggregated over a time period according to the type of queries by individual officers and 3) the computer communication differences between a cellular digital packet data system and proprietary Motorola system are compared.

Types of queries are defined as queries that access different databases or different primary fields within multiple databases to provide information about a person, article, vehicle or situation. The multiplicity of cues in this study refers to the different types of information (*i.e.* when looking at vehicles officers might want to know the registered owner, crashes, previous insurance claims, stolen status, or if the vehicle was associated with previous crimes). This additional data presents the officer with a richer base of information to formulate a decision.

The 1994 Naturalistic Decision Making (NDM) conference defined the study of NDM as "how experienced people, working as individuals or groups in dynamic, uncertain and often fast-paced environments, identify and assess the situations, make decisions and take actions whose consequences are meaningful to them and to the larger organization in which they operate." This definition was expanded by Orasanu and Connolly (1993) to eight characteristics (Table 1).

Table 1: Naturalistic decision model characteristics

	NDM Characteristic	**Characteristic match in research setting**
1	Ill-structured problems	Always incomplete and equivocal information
2	Uncertain, dynamic environments	Adversarial situations in a wide variety of settings
3	Shifting, ill-defined or competing goals	Immediate goals are defined by each situation and shift between personal safety, individual rights, public safety and societal perception
4	Action-feedback loops	Initial decisions reinforced by more data to clarify incomplete and conflicting data
5	Time stress	Legally imposed time limits and standards
6	High stakes	Serious injury or death, freedom or incarceration, & lawsuits for improper action
7	Multiple players	Decision-makers facing new people every day
8	Organizational goals and norms	Standard practices and policy influenced by changing legal standards

All eight elements of naturalistic decision making above (Orasanu and Connolly 1993) are described in the research setting portion of this paper. The characteristics in my study match best only one of the five (5) major naturalistic decision models, the recognition-primed decision model (Zsambok 1997).

Scope of the Study

Archived, empirical data gathered from a period covering six (6) years is used to evaluate the research model. These data describe the deployment of a large scale, wireless computer system used by a city-county police department. Mobile police officers and the resulting arrest records are scrutinized over a six-year period to evaluate the impact on arrest rates of varying media richness (from voice mediated radio access to self-initiated queries to retrieve data through textural and verbal query interventions) using computers of varying wireless query, access-retrieval, rates. Outcome measures consist of arrests made and warrants served, where officers attempt to identify a person named in a legal document and place them in custody, plus physical arrests based on probable cause. It is important to note that all persons placed in custody, for whatever reason, are immediately taken before a neutral third person, a magistrate, who would assess whether the information and/or circumstances reach a decision threshold of probable cause; that this person is the person named in the warrant or is the person who committed a crime. Negative determinations of probable cause, where a magistrate rejects a warrant or arrest are rare to the point that these rejections are not recorded as arrests in a person's record or officer's performance.

All outcome measures are observed pre and post interventions. The aggregate performance of the officer, by time period, is the unit of analysis.

Interventions

There are three interventions (historical deployments of wireless computers):

Base Level - as given in example case one

Prior to the deployment of computers, officers would obtain information from local, state and federal databases by requesting information to support or refute a particular subject's guilt or innocence (according to the face implications as stated on the warrant for arrest). They were able to access this information by radio by having a third party query the information at a wired computer terminal and read it back to the officer over the radio. Normalized pre-computer base levels of activity by field police officers were measured that included: numbers of arrests per day and the number of warrants served per day. These tasks were supported using a radio by the officer to contact another person who would run queries on a remote computer terminal and report query responses back to the officer, orally by radio. Officers are required to use the technology. Numbers of queries per day and numbers of types of queries used per period are collected in large numbers of similar decisions in la.

Control is through use of pre and post officer performance outcome measure observations (*i.e.* these outcome measures are how police measure field officer performance). The pre-intervention data is gathered through the performance of individual officers prior to any field computers being in use and data normalized for seasonal, temporal or geographical variations. Multiple observations between interventions make a stronger argument in any changes thought to be due to information latency. The total arrests line displays general arrest trends (see Figure 10).

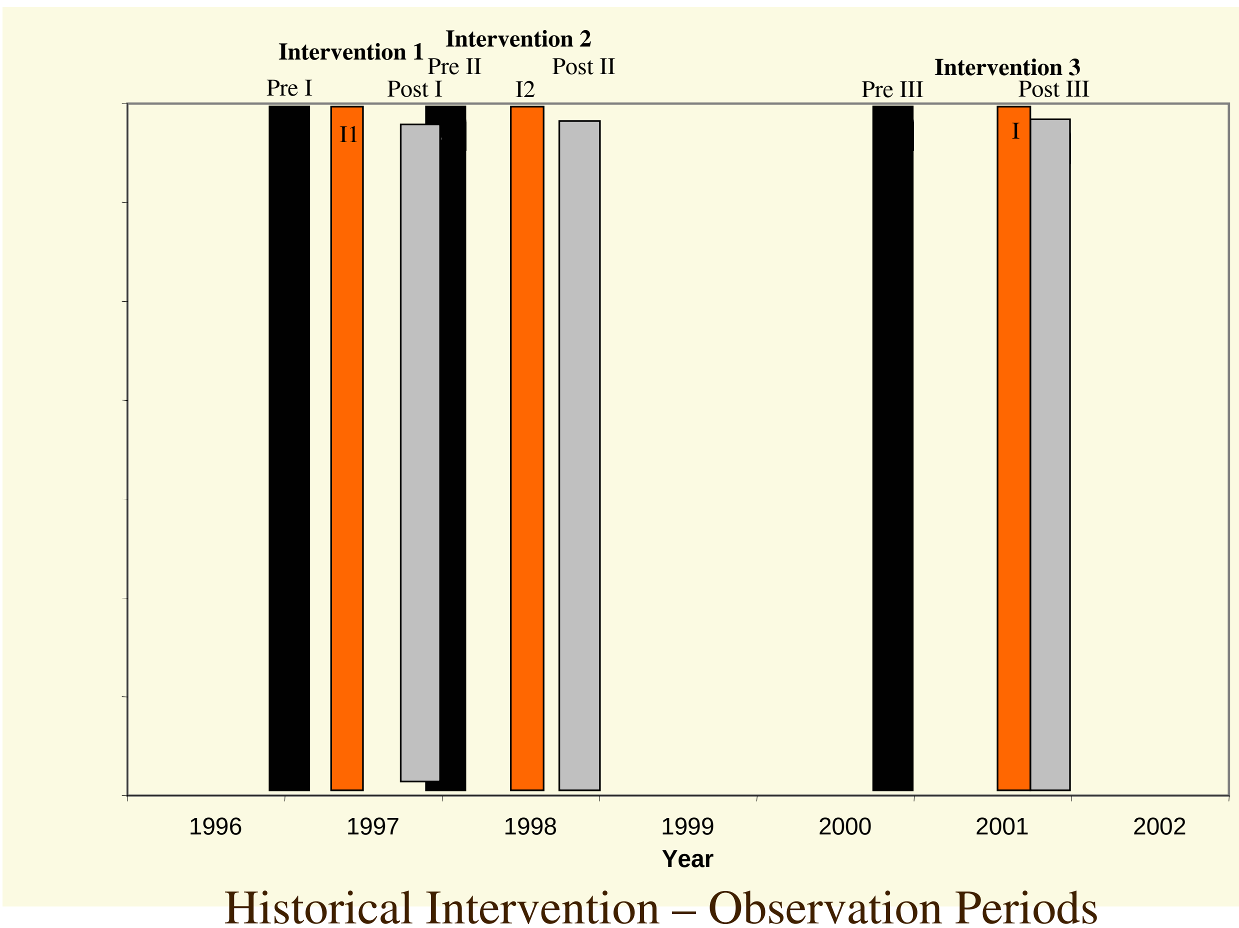

Figure 10: Three Pre-Post Observations of Historical Treatments

Historical Intervention I (Treatment I): lowered information latency as given in example case two

Wireless computers were then issued to a select group of officers in 1997 (Group B) to allow self-initiated textural queries to replace the radio-voice retrieval method. Those without computers are Group A. The performance of officers was measured as the mean number of warrants served and mean number of arrests made. This historical intervention (treatment) replaces a third party running a query at a distant source and verbally reading a response back to the officer. A key change with this treatment meant officers no longer had to wait in a queue with other officers throughout the police department to verbally tell the remote clerk what to run. Speaking a seventeen digit, alpha-numeric vehicle identification numbers alone would take considerable time and increase a likelihood of a transcription error when the listener misinterprets a "B" or "V" for example.

Historical Intervention II (Treatment II): lowered information latency (example case two)

Wireless computers were issued in 1998 to another group of officers (Group C, a subset of Group A) in patrol who had been only using verbal radio communications for queries (spread evenly geographically across the county). These computers were identical in communication and query speeds with computers from intervention one (Group B). Those officers still without computers are in Group D (a subset of Group A also).

Historical Intervention III (Treatment III - Group E): higher information latency as given in example case three

Wireless computers with moderate latency, slower access speeds, than in Intervention I or II were issued to a group of officers (Group E) who had not been issued the lower latency computers. Group E is essentially those remaining from Group D in Intervention II. This intervention uses new software that requires moving larger files over a slower network but having faster internal processing speed. Overall information latency is higher than Intervention I or II but significantly lower information latency than radio queries. This moderate increase in information latency is still statistically significant (see information latency section).

(Group F): higher information latency as given in example case three

Higher latency computers were given to a group of officers (Group F) who had been using lower latency computers. This group is a subset of group C and they began using the new wireless computers from Intervention III to replace the computers from Intervention II. There is increased information latency when moving from computers in Intervention II to Intervention III (statistically significantly increased from intervention I and II – see information latency section).

(Group G): moderate information latency as given in example case three

This group also is a subset of Group C uses wireless computers with the access speed from Intervention II but new software is added that requires moving a larger file. Overall latency is slightly higher than during Intervention I and II.

(Group H): moderate information latency as given in example case three

This group uses the same wireless computers as in intervention III because it replaces computers from Intervention I during the same time period as Intervention III. There is a moderate change in latency when moving from computers in Intervention I to Intervention III.

Population and Sample

The unit of analysis is the user and decision-maker (the field police officer) making repeated high-risk decisions. The individual users' performance outcomes were used pre and post treatment, similarly to the Task-Technology Fit model (Goodhue and Thompson 1995), to evaluate changes caused by the treatments. Performance outcomes, in this research are the average number of warrants served per day or the arrests made per day by *each individual officer*. Officers are required to use this technology so acceptance and utilization is not a factor. The sample population is taken from street-level police officers who make life-death, arrest-freedom decisions on a daily basis in the course of their work.

The Metro Nashville Police Department implemented a wireless laptop system in 1997 with 34 uniformed police officers who had never used laptops previously. One hundred and twenty (129) new users were added in 1998. Three hundred new computers were deployed in June through August 2001. Subjects were followed from their pre-computer, 1996 stage, through each intervention (when information attributes changed dramatically) until December 31, 2001. There are 406 in the total sample.

There are approximately thirty-four (34) officers for the first treatment design (Intervention I).

NR 0 X1 0 X3 0

There are approximately one hundred-ninety-two (192) officers for the second treatment design (Intervention II).

NR 0 X2 0 X3 0

There are approximately three hundred four (304) in the third treatment (Intervention III).

NR 0 0 X3 0

Information Latency:

I = II < III < Base Case

Where the information latency with new computers in Intervention I and II are the same, increased latency with intervention three and new computer latency is lower than radio base case.

Group Definitions and Historical Interventions (treatments)

There are three periods when computers were issued (*i.e.*, three historical interventions) within the study period. These group definitions, interventions, and periods are used as the basis for pre and post comparisons and occurred during the following dates (see Figure 11 & Table 2):

GROUPS AND INTERVENTIONS

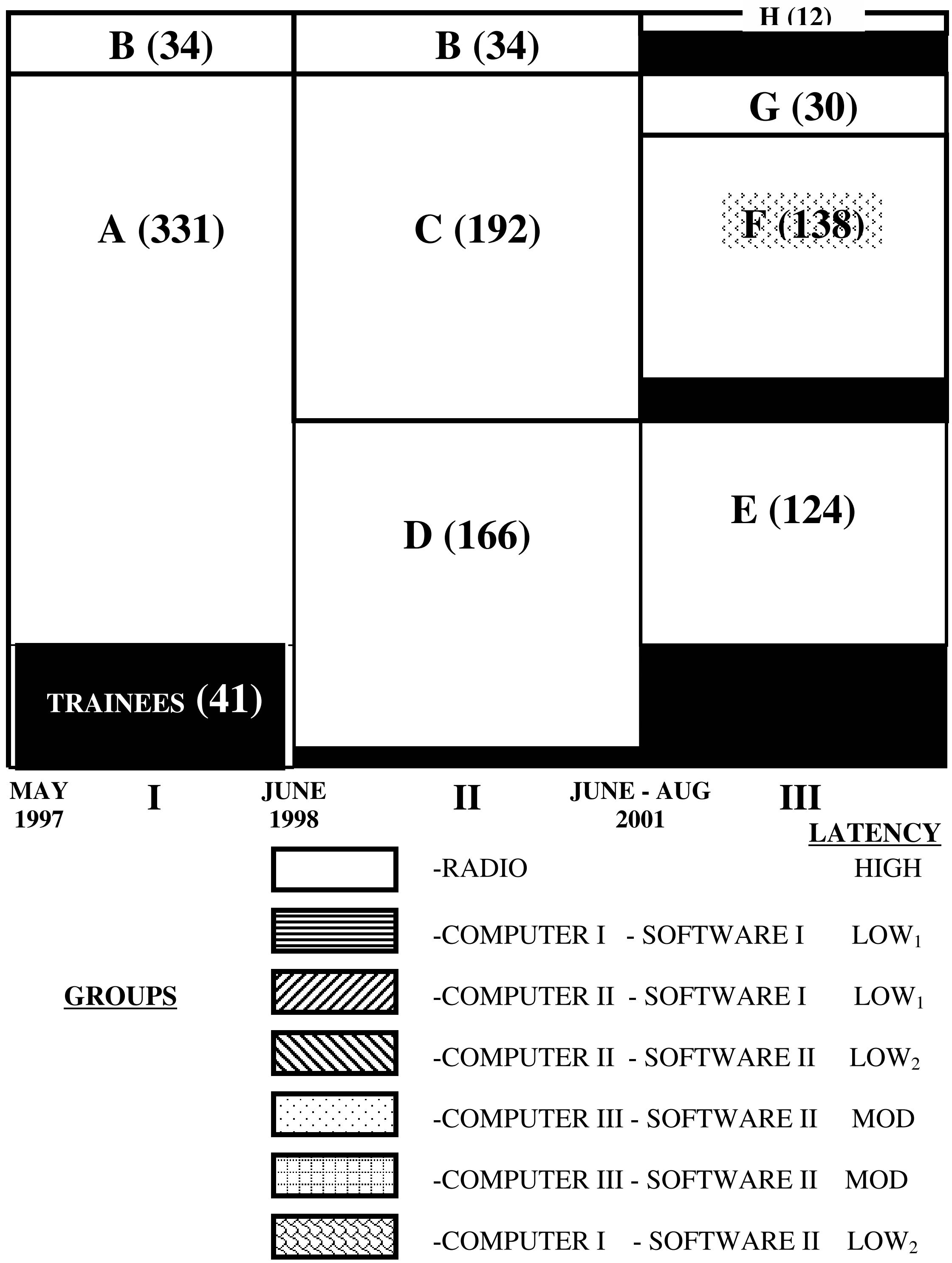

Figure 11: Interventions and Groups

Table 2: Intervention Groups, Time Periods and Counts (% of sample)

Intervention	Pre-Test Period	Treatment Period	Post-Test Period	(%) Of N	Description
I (Group A)	12/1/96 to 2/28/97	March – May 1997	12/1/97 to 2/28/98	365 (90%)	331 without
(Group B)	”	”	”	”	34 new users
II (Group B)	10/1/97 to 12/31/97	June – August 1998	10/1/98 to 12/31/98	392 (97%)	34 from I
(Group C)	”	”	”	”	192 new users
(Group D)	”	”	”	”	166 without
III (Group E)	10/1/00 to 12/31/00	June – August 2001	10/1/01 to 12/31/01	304 (75%)	124 new users
(Group F)	”	”	”		138 replacing II
(Group G)	”	”	”		30 upgrading II
(Group H)	”	”	”		12 replacing I

Independent Variable

The independent variable in this research is information latency, in this case, the time from query to receipt of information. Also the information latency differences are aggregated using the most conservative mean query to receipt/display times according to the each treatment. See information latency tables in the analysis section.

Dependent Variables

The dependent variables of interest are the mean number of arrests and mean number warrants served by each officer within a specified time period (outcome measure of the decisions). The warrant description and the computer textural data description are the same in this case (warrants are computerized). The officers must determine from uncertain, equivocal information, using textural cues to determine if a person is, or is not the subject of the warrant (a binary – yes/no decision). The fact building with warrants was to 1) confirm the identity of the person being questioned and 2) this fact building is indifferent to crime elements[8] at this point in

[8] ***Elements of a crime*** are *the components that make up a criminal law violation (i.e. burglary consists of 1) a breaking and entering 2) with the intent to commit a felony).*

the field because a legal document to arrest has already been issued and identity is the main consideration.

Arrests made without a warrant are based on probable cause (where a reasonable person would believe that this person had committed a crime). Facts are gathered by an officer in the field to reach a "threshold" of belief that a person has committed a crime. This fact building is to 1) identify the elements of a crime and 2) it is indifferent to the identity of the person other than to connect this person to the crime that is purported to have been committed. Then the arrestee is taken before a magistrate to act as an independent third party to rule whether there are sufficient grounds (facts) to issue a warrant. I examined changes in performance outcome measures (number of arrests and warrants served).

Mediating Variables

The mediating variables of interest are feedback, (*i.e.* the number of queries per officer in designated periods of time), and multiple cues (*i.e.* the number of types of queries). The pre-intervention data were recorded as a baseline for each study subject by showing the areas worked, number of previous warrants served/arrests during pre-computer stage and the times required to retrieve data in a voice response over a radio. The database used for queries has regularly updated data that is timely (current), and by state and federal regulations, tightly regulated for quality using only certified operators (regulated by each state). Types of queries are textural queries that access different databases, different fields within databases or different combinations of information about people, objects/property, vehicles or places. Types of queries were considered as multiple cues, for this research.

Control group

Threats to validity were evaluated because of the secondary data used. Some officers who had computer systems with low information latency in 1998 were required to obtain new software in 2001. This software required them to move more data (larger files) with the same throughput access speed. A separate set of officers who had never had a computer obtained computers that were set up to use a slower access communication that, due to a capacitated network, would only operate at a fraction of the optimal access speed. Switching replications increased the likelihood that any changes are not a result of threats to validity. Some of the study participants (officers) in the first two interventions (approximately 150) received a replacement computer of different data access speed. All the study participants received short, two to three hour, training with their laptops. There were no advanced classes for any study participants. The similar training for groups with computers removes differences in training as a variable. The short period between receiving a computer and post-intervention observation reduces maturation threats. Temporal precedence is reduced with multiple observations during multiple treatments.

Due to the six year length of this study; there are other threats to validity that must be countered. Two of the largest threats were to adjust for variations in performance (arrest rate and warrant service rate adjusted for seasonal variations) and mortality. These threats were mitigated by paired, multiple tests of performance and having a large enough sample size throughout the research. Officers who lost laptops were eliminated as a control throughout my research to counter history threats. Performance variables were normalized to each participant to counter maturation effects. Police officers perform many tasks (i.e. directing traffic, writing reports, checking building, giving directions, assisting stranded motorists, etc.) but the tasks in my research are normalized to the tasks of making arrests or serving warrants.

CHAPTER IV

DATA AND RESULTS

Sample Population

There were 1,191 sworn police officers in the study population during the period of the first computer issuance in 1997. Of that total, there were approximately five-hundred (500) officers assigned to uniform patrol duties under the rank of captain in the sampling frame. In the period of 1997 through 2001, the minimum staffing level in patrol (actually answering calls for service) was set at 420 officers. The number "five hundred" is a misleading number as there are normally 18 to 20% of the officers on special assignments, on vacation, at in-service training or out sick or injured. At any one time, there are approximately 78 officers, actually on-duty in a shift. These are eight-hour shifts, seven days a week, 24 hours a day.

The sample consists of 406 officers identified as being assigned to patrol (uniformed) (96.6 % of the 420 – minimum staffing) street functions, during the period where the archival data was examined, that used a wireless computer during any of the three historical treatments (when computers were issued). These officers and their data were identified through electronic records of issuance and through printouts of their query aggregates per month.

Sources of the data

Officer demographic data comes from personnel rosters for the officers during this seven (7) year period from 1996 to 2002. Rosters are used to indicate whenever an officer has left police service so that officer can be removed from the study. Arrest and warrant data come from the archived six hundred (600) daily field activity summary reports. The Uniform Services Bureau of the police department compiled these reports over the 2,556 days studied. Each segment of performance outcome measures includes eighteen categories of action in spreadsheet form in monthly segments. Officers were tracked using employee numbers for subject identification over the years of the study. Numbers and types of queries per officer come from mainframe usage data and middleware databases in monthly paper rosters for the first five years of the study. For the next two years of the study, these results come from digital, delimited text files.

Pre-computer radio query averages for numbers and types of queries were provided by the Police Planning and Research Division from a 1997 evaluation of a "COPS More 96" grant. Radio and computer response times came from field recording of data by personnel in the police Planning and Research Division who were capturing data for an evaluation of the computer deployment in 1997. The radio query data was captured from 30 samples monitored over the police radio system.

Police radioed query totals, from all officers through clerks, were gathered using log scans of the fixed remote terminals used by these clerks. The log scans list total radio queries for all police by monthly segments starting in December 1998. These queries are referred to as *radio queries*. The information services division of the police department provided dates of computer issuance, upgrade, and service periods that designate when different computer configurations are in (or out) of use during the study period.

Total arrests per day and total number of warrants served are collected daily by supervisors and aggregate totals are used in performance appraisals. Police supervisors, as part of their daily activity, consistently record outcome arrests, warrants served and days worked by their officers throughout the study period.

Data Clean-up

Inspection of the mean arrests per day and mean warrants served per day data per officer demonstrated that outcome measures were exponentially distributed and needed a log transform to make them linear for statistical evaluation. Those few data outliers falling outside twice the standard deviation for the variable were trimmed before ANOVA comparison of means among groups. This trimming typically occurred among those officers that were very high performers. The following data points were trimmed: five (5) out of 365 mean log arrests per day pre-Intervention I and five (5) post Intervention I. Two (2) data points out of 365 mean log warrants served per day pre-Intervention I and one (1) post, two (2) data points out of 392 mean log arrests per day post Intervention II, three (3) data points out of 304 log mean warrants served per day pre-Intervention III with one (1) post, one (1) log mean arrests per day and two (2) log mean queries per day post Intervention III. Examples of outliers removed in this last case are where the two data points were 6.10 and 5.212 mean log queries per day, both well above $2(\sigma)$ + mean (where $2(\sigma)$ + mean = 4.7885). Figure 12, below, shows the two outliers for this variable.

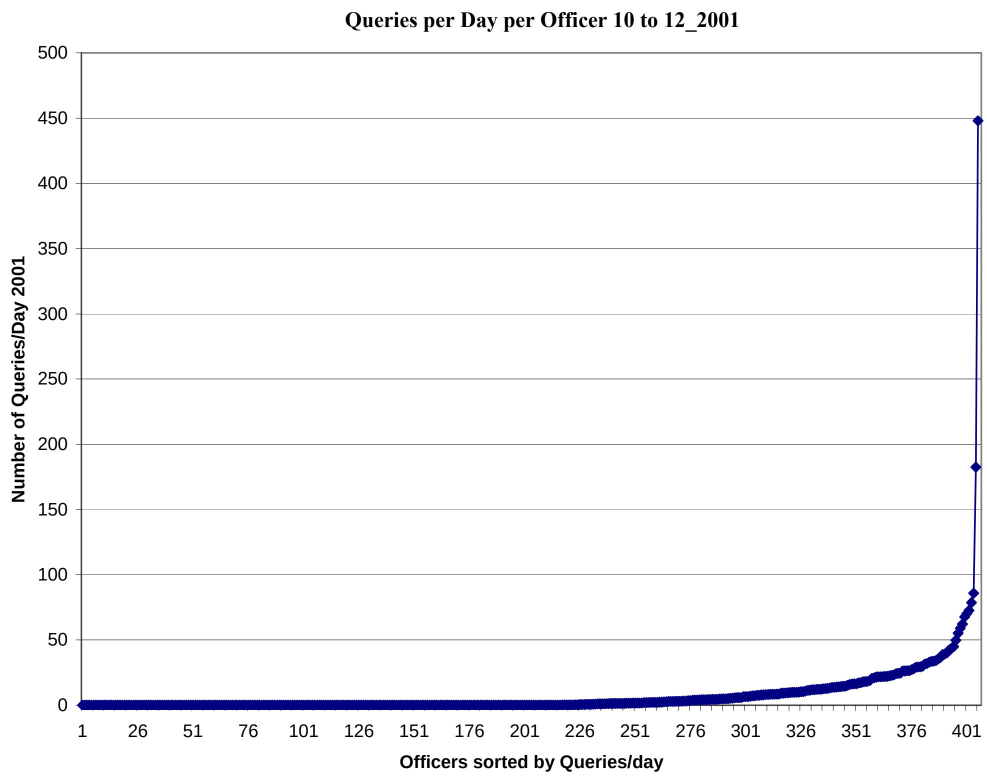

Figure 12: Removing outliers from queries per day

Statistical data analysis

The model is tested first by determining levels of information latency during each historical intervention period in the study. This includes information latency for radio queries, pre-intervention to information latency changes when any new computer system was introduced, upgraded, or replaced. Each of the three hypotheses and subsets are then tested across interventions, where data is available. User initiated query responses to the decision-makers are evaluated for their information latency and media richness characteristics. The differences in query information latency (transmission-to-retrieval-to-display speed) and using radio information latency (radio-voice-third party method) were compared as was wireless textural information access of different speeds. The alpha value for this study is 0.01 (α = **0.01**) to reduce likelihood of Type I errors.

Two types of evaluation are completed *within* the groups. First, paired t-tests are used to examine differences between individual performance before and after each intervention for each hypothesis if the data is available. Second, secondary analysis, comparison of two means, through significance tests on the confidence interval is used to examine mean radio queries/day versus post Intervention I queries and for pre-Intervention III radioed mean queries/day and mean types of queries.

Testing within groups

For Hypothesis 1, the primary outcome measures are mean queries/day and mean types of queries per month. In Hypothesis 1a, the specific statistical analyses performed are: paired t-tests by officer performance for Intervention III, using outcome measures of mean queries/day. The reason complete pre- and post- intervention analysis was not completed for Intervention I and II is because aggregate wireless query data was not available until post Intervention II (December 1998). However, data was available for a 166 day period for officers either with or acquiring computers during Intervention I. Using these data, it is possible to estimate the possible impact of Intervention I on queries per day by officer using comparison of means pre- and post Intervention I through comparison of the differences in sample proportions and means.

Statistical analysis performed for Hypothesis1b is paired t-tests for pre- and post-Intervention III groups by types of queries. Secondary analysis is performed, by large sample significance test, for Group E by using estimated mean types of radio queries pre-intervention III. As with H_1a, the reason complete pre- and post-intervention analysis was not performed for all interventions is because data were only available for periods December 1998 to the end of the study. Mean types of queries, by radio, comes from a 1998 estimate by the police planning and research division.

For Hypothesis 2, the primary outcome measures are mean arrests/day and warrants served/day. In Hypothesis 2a, the specific statistical analyses performed are: paired t-tests for Intervention I, II, and III. Also paired t-tests were performed across the groups in Intervention III by queries/day. The reason complete pre- and post- intervention analysis was not completed for all interventions are because aggregate wireless query data was not available until post Intervention II.

Statistical analyses performed for Hypothesis2b are paired t-tests for pre- and post-Intervention I, II, and III groups with mean arrests/day and mean warrants served/day. The reason complete pre- and post-intervention analysis was not performed for all interventions is because data were only available for post Intervention II.

Hypothesis 2C combines the statistical analyses performed for Hypothesis 2a and 2b, paired t-tests of Intervention I, II, and III mean arrests per day and mean warrants served/day. Paired-t-tests of mean queries/day and types of queries were performed pre- and post-Intervention III.

Hypothesis 3 is examined using the previous paired t-test analyses of mean arrests/day, warrants served/day, and queries/day during Intervention III where the three groups experienced, respectively; large, moderate or small changes in information latency. Secondary analysis by comparison of means by queries per day by officer using comparison of means pre- and post Intervention I and II is performed through a significance test of confidence intervals.

Information Latency

Aggregate information latency data was gathered from a number of sources. Data collected by the police planning and research division (30 samples) in 1998 revealed that radio queries by patrol officers took over 2 minutes, after the officer waited up to twenty minutes in a virtual queue (mean queuing times were estimated at 5 minutes), until the sole remote operator could take their information over the radio transceiver and run the query from a distant terminal (data from minute logs). Mean radio latency was added to the queuing time to obtain the estimated radio latency. Queuing latency was twice the actual radio query latency in most cases. The time the query was broadcast by the officer over the radio until the verbal response back to the officer, was recorded to the second by personnel from the Planning and Research Division. (Table 3):

Table 3: Information Latency for Different Interventions

Type Query	Mean Radio Latency 5 Min Queuing (1998) + Radio Query = Total (Very High) Latency				Mean Latency Int. I, II (1998) (Low)
	Queue	**Radio**	**Total**	**n**	
Vehicle Registration	5:00	1:58	6:58	94	0:02 (n=30)
Criminal History	5:00	3:50	8:50	188	0:04 (n=30)
Driver's License (State)	5:00	2:56	7:56	57	0:12 (n=30)
Warrant Check	5:00	2:30	7:30	175	0:04 (n=30)

Table 3 (above) shows sampled query times for different types of queries for the first two Interventions.

Statistical significance of information latency changes across historical interventions

Information latency is not available in individualized data for every query by every officer but average query response times are available for four types of queries. The aggregate means of these queries are compared using large sample comparison (Agresti and Finlay 1997) using the most conservative times (*i.e.* the least mean time for any radio query versus the most time for *any* type of wireless query – arrows denote the change in information latency across the intervention where ↓ is a reduction in information latency. Arrows where ↔ indicate pre and post technology are the same). Radio without queuing is if officers without computers radioed queries to officers with computers (see Table 4 below).

Table 4: Information latency changes within groups across interventions

Group Identification	Pre-treatment	Post-treatment	Sign of mean differences (p-value)
I (Group A)	Est. Radio with queuing	Est. Radio with queuing	↔1.000
(Group A)	Est. Radio with queuing	Est. Radio without queuing	**↓0.000**
(Group B)	Est. Radio with queuing	Wireless queries	**↓0.000**
II (Group B)	Wireless queries	Wireless Queries	↔1.000
(Group C)	Est. Radio with queuing	Wireless queries	**↓0.000**
(Group C)	Est. Radio w.o. queuing	Wireless queries	**↓0.000**
(Group D)	Est. Radio with queuing	Est. Radio with queuing	↔1.000
III (Group E)	Est. Radio with queuing	Slower wireless queries	**↓0.000**
(Group F)	Wireless queries	Slower wireless queries	**↑0.0012**
(Group G)	Wireless queries	Slightly slower wireless	Not available
(Group H)	Wireless queries	Slower wireless queries	**↑0.0012**

Changes in information latency were calculated using the times and counts from Table 4 with the most conservative differences between times to demonstrate that the least difference between any types of queries is statistically significant or unchanging using pre and post treatment averages. Using the least differences between queries allows the differences between any query types pre- treatment to be compared to any query type post – treatment. Estimated aggregate times for radio queries came from 1997 and 1998 estimates of timed queries (police planning and research division – 30 samples). Radio queries *without queuing* are radioed queries to other officers who would run the queries on their wireless laptops.

Further explanation of query types

Types of queries are defined as queries that access different databases or different primary fields within multiple databases to provide information about a person, article, vehicle or situation. There were 76 different types of queries, identified from printouts of individual users, being used in December of 2001 and this number of query types was available in 1997. The following are typical query type examples (not inclusive): criminal warrants, criminal history, vehicle registration, juvenile history, nicknames, demographics, jail management, driver's license, stolen property, traffic accidents and similar. Officers are free to use any queries they feel will give them information to clarify anomalies. Some queries refer to vehicles, some to persons and identities or history and others refer to property or past incidents. All queries are not applicable in all situations or officers would be able to simply ask a standard set of questions in a set order.

Detailed records of types of queries by officer, aggregated by month, do not exist until post-Intervention II. Comparisons of aggregate query records of all wireless users after December 1998 to the end of the study period show that officers performing wireless queries use many of these 76 types of queries compared to the four typical radio queries (warrant check, arrest offense records, criminal history, or vehicle registration) for officers radioing queries. These queries will be referred as the *basic four* queries. The basic four queries account for approximately 90 per cent of radioed queries and also account for the majority of queries for wireless users: 71% of the queries in 1998 (Oct. – Dec.), 75% in 1999 and 79.7% in 2001 during comparable periods (see Table 5).

Table 5: Top Ten Types of Queries and Counts December 2001

Description Of Type Of Query	**Number Per Type**	**%**
Check For Active Warrants On File	27,108	**30.9%**
Arrest Check By Name, Driver's License, Soc Security Number	19,797	**22.6%**
Criminal History Check	11,562	**13.2%**
Vehicle Registration Check + Stolen Local And State	11,360	**13.0%**
Local Check For Orders Of Protection, Citations	2,328	2.7%
Local Motor Vehicle Registration Information	1,877	2.1%
Queries Of The Employee Records File	1,685	1.9%
Extra Id Query	1,474	1.7%
In-depth Criminal History I	1,095	1.2%
Juvenile Arrest History	1,043	1.2%
…And 55 Other Types Of Queries: Jail Management, Zip Codes, Associated Names (Court, Witnesses, Codefendants), Court Dates, Probation, Gun Applicants, Gun File, Etc.	8,312	9.5%
Total	87,641	100.0%

Intervention I – March to May (1997)

The first test introduced wireless computers in Intervention I. User query records reported 34 uniformed patrol officers performing wireless computer queries (Group B). Early records, 1997-1998 only showed total queries over dissimilar length, daily periods (non-comparable periods) with no reporting on types of queries. (An example compares Officer X, who worked from May 16 to May 29th, 1997 and performed 673 wireless queries to Officer Y, who used another, similar computer from June 16th to June 23rd, 1997 and performed 1,016 wireless queries). The activities (arrests and warrants) of December 1st, 1996 through February 28th, 1997 and December 1st, 1997 through February 28th, 1998 are used for pre and post-periods for intervention one. The data is gathered for 331 officers without computers (Control Group A) and the 34 officers identified as uniformed (patrol) officers (Group B) who were using the wireless laptops (n= 365). There were 41 officers who were trainees during Intervention I and neither had radios or wireless computers.

Table 6: Intervention I Groups, Time Periods and Counts (% sample)

Intervention	**Pre-Test Period**	**Treatment Period**	**Post-Test Period**	**(%) Of N**	**Description**
I (Group A)	12/1/96 to 2/28/97	March – May 1997	12/1/97 to 2/28/98	365 (90%)	331 without
(Group B)	”	”	”	”	34 new users

GROUPS IN INTERVENTION I

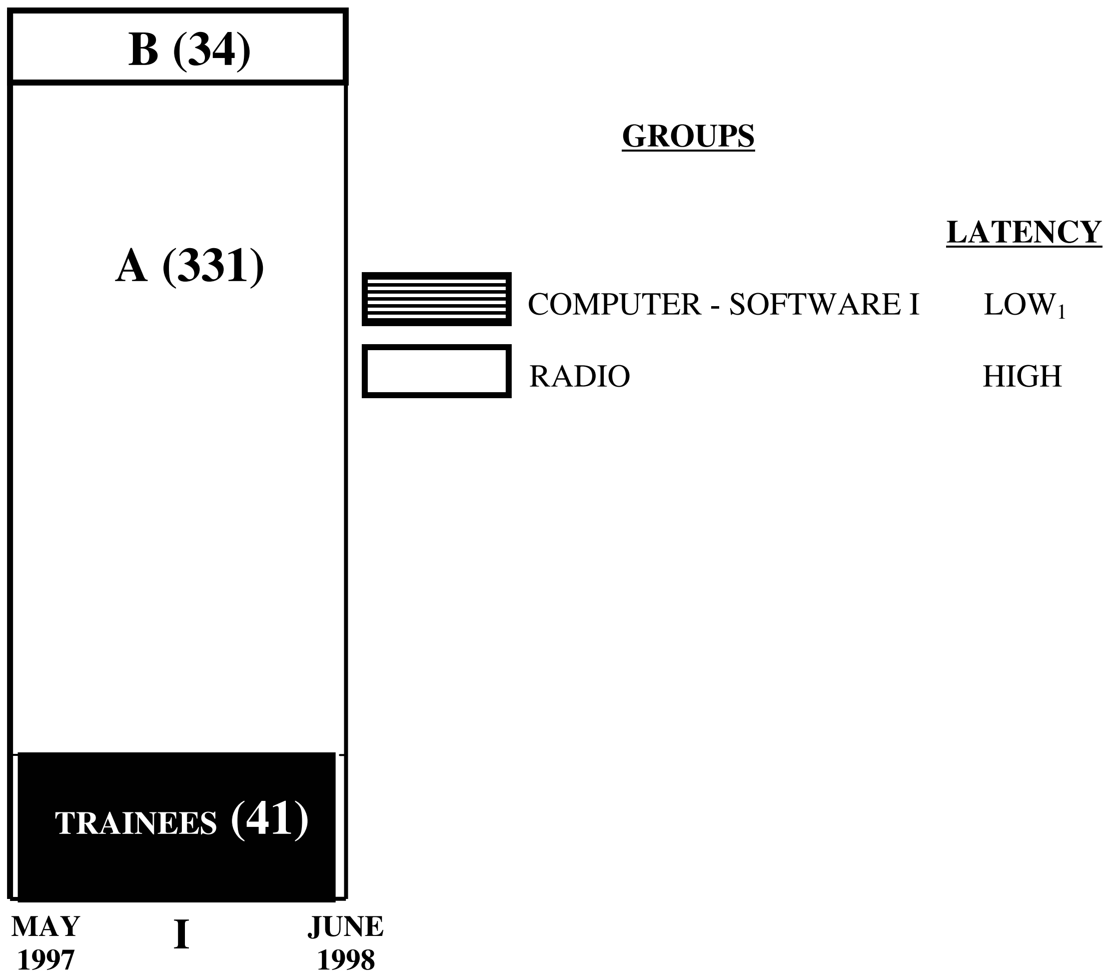

Figure 13: Groups in Intervention I

Table 7: Available Data and Source

Intervention	Group	Pre- or Post- Data	Measure	Source/Availability
I	A, B	Pre-	Mean radio query times (latency)	1997 Planning and Research survey (no standard deviation)
	A, B	Post-	Mean radio query times And mean computer I query times (latency)	1997 Planning and Research survey (no standard deviation)
	A, B	Pre-	Mean radio queries/day	1997 Planning and Research survey (no standard deviation)
	A	Post-	Mean queries/day	Not Available
	B	Post-	Mean queries/day	Number Queries and estimated Days Worked
	A,B	Pre-	Mean types of queries	Not Available
	A,B	Post-	Mean types of queries	Not Available
	A,B	Pre-	Mean Arrests/day	Activity reports
	A,B	Post-	Mean Arrests/day	Activity reports
	A,B	Pre-	Mean warrants served/day	Activity reports
	A,B	Post-	Mean warrants served/day	Activity reports

Wireless queries identified each user of any wireless computer but did not show monthly intervals until later in the study. Wireless computer query transactions totals are referred to as *wireless query totals*.

The computers introduced in Intervention I ran a 3270 terminal emulation application that supported all query types (equivalent in information latency to the terminal emulation application used on remote terminals by records clerks according to tests by the Police Planning and Research Section May – June 1997).

Secondary analysis by comparison of means by queries per day by officer using comparison of means pre- and post Intervention I is performed through a significance test of confidence intervals. These tests are listed in Table 8 (below).

Table 8: Statistical Tests of Hypotheses within Intervention I Groups

Hypothesis tested	**Measures used**	**Statistical tests**
Hypothesis 1a	Mean queries/day	Two Mean Comparison
Hypothesis 2a	Mean Arrests/day	Paired t-test
	Mean warrants served/day	Paired t-test
Hypothesis 2b	Mean warrants served/day	Paired t-test
	Mean Arrests/day	Paired t-test
Hypothesis 2c	Mean warrants served/day	Paired t-test
	Mean queries/day	Two Mean Comparison
Hypothesis 3	Previous tests- 2a,b and c	Previous tests

Hypothesis$_{1A}$ testing for Intervention I

Hypothesis1$_A$ states that information latency affects feedback in high-risk decision-making. The reduction in query latency should increase the number of queries for the officers across the interventions regardless of query type. Pre- and post- data by mean queries per day is not readily available for Intervention I. There were records on total queries for one 166 day period post Intervention I for Group B as will be described in a secondary analysis that follows as we examine information latency on the model (Figure 14 and Table 9).

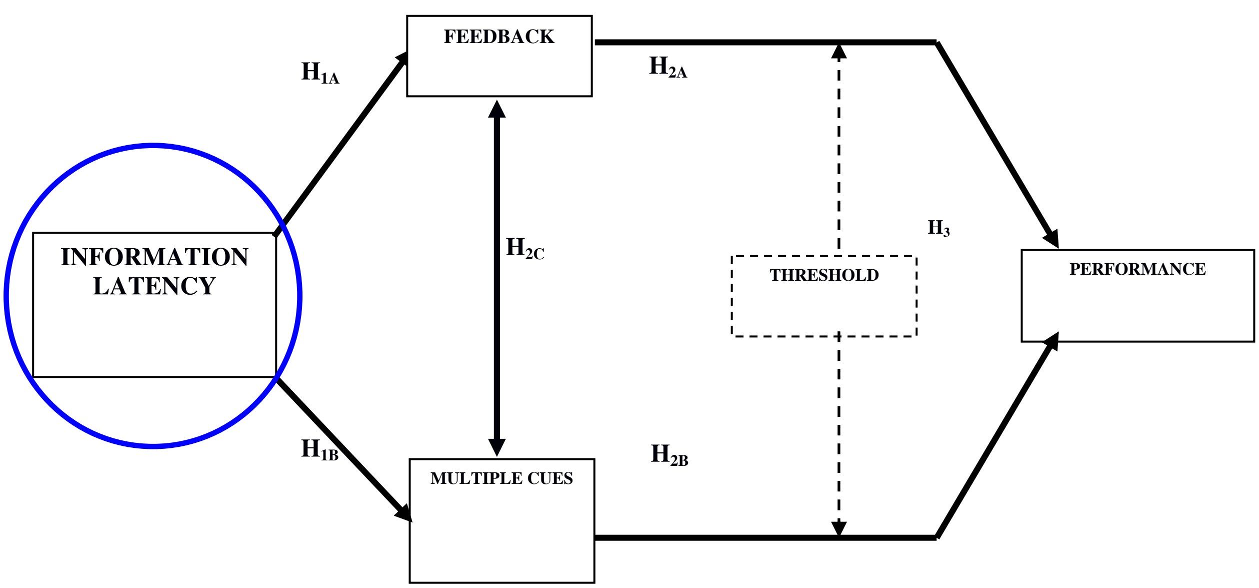

Figure 14: Testing the information latency portion of model

Table 9: Statistical Comparisons of Means for Information Latency Change (Group aggregates)

Group Identification	**Pre-treatment**	**Post-treatment**	**Δ Comparison of means (p)**
I (Group A)	Est. Radio with queuing	Est. Radio with queuing	↔1.000
(Group A)	Est. Radio with queuing	Est. Radio without queuing	↓**0.000**
(Group B)	Est. Radio with queuing	Wireless queries	↓**0.000**

Secondary analysis of mean number of queries with Intervention I Group B

While pre- and post mean queries/day data were not available until post- Intervention II, the record of wireless queries was available showing the number of queries generated by each of the 34 officers in Intervention I (Group B). That is why the secondary analysis only uses Group B. This data shows Group B officers generating 79,008 queries in the 166 days of April 25th through October 7, 1997 (during treatment and post Intervention I). However, officers would not have worked more than 22 days per month (5 day work week) during the 5.5 months in that period, meaning they would have worked no more than 121 days out of the 166 days available (the most conservative estimate of possible days worked). The mean queries per day are then estimated to be (79,008 / (34 officers * 121 days worked)) equal to 19.2 queries per officer per day. The pre-computer estimate is six (6) radio queries per officer per day. Using mean comparisons, with the estimated standard deviation for the typical radio query per officer as equal to the standard deviation of the computer Intervention I officers, we can compute the z-score using the following formula (Agresti and Finlay 1997):

$$\hat{\sigma}\overline{Y}_2 - \overline{Y}_1 = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\hat{\sigma}\overline{Y}_2 - \overline{Y}_1 = \sqrt{\frac{19.5^2}{420} + \frac{18.5^2}{34}}$$

$$\hat{\sigma}\overline{Y}_2 - \overline{Y}_1 = 3.31$$

$$z = \frac{\overline{Y}_2 - \overline{Y}_1}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$z = \frac{19.2 - 6}{3.31} = 3.98$$

$$p \cong 0.000$$

There was a statistically significant increase in mean number of queries per day with a change in latency with Group B for Intervention I (Figure 15).

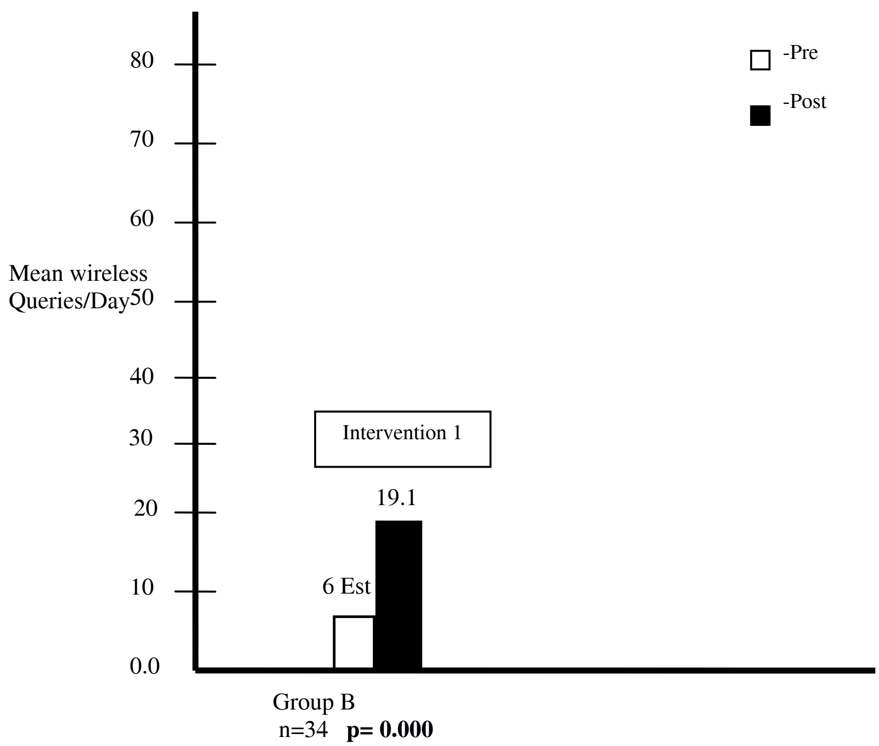

Figure 15: Mean Queries per day for Group B

Hypothesis$_{1A}$ is supported in Intervention I that information latency affects feedback in high-risk decision-making. Statistically significant changes in latency in Group B resulted in statistically significant (p=**0.000**) changes in the mean number of queries per day.

Hypothesis1_B testing

Hypothesis1_b states that *information latency affects multiple cues in high-risk decision-making*. The change in query latency is expected to change the number of types of queries across the interventions. Paired t-tests could not be performed on intervention I and II by mean type of queries because mean types of queries data was not available until post- Intervention II.

Hypothesis$_{2A}$ testing

Hypothesis$_{2A}$ states that *feedback*, measured *by the mean number of queries per day, affects high-risk decision performance.*

It is expected an increase in the mean number of queries per day allows more clarification of anomalies in high-risk decision-making. Data from each of the three interventions should demonstrate a change in the performance outcome measures across each intervention for the computers users with changes in mean number of queries per day.

Intervention I outcome comparison within groups

The 334 officers in Group A, with no computer, show statistically significant increases in both mean arrest rate/day (p=**0.000**) and mean warrant service/day (p=**0.000**) in paired t-tests across Intervention I. Note: This is opposite of the expected direction. The t-test for those in Group B, with fast wireless, showed no statistically significant change in either mean arrests/day (p=0.340) or mean warrants served/day (p=0.915). (Figures 16 and 17).

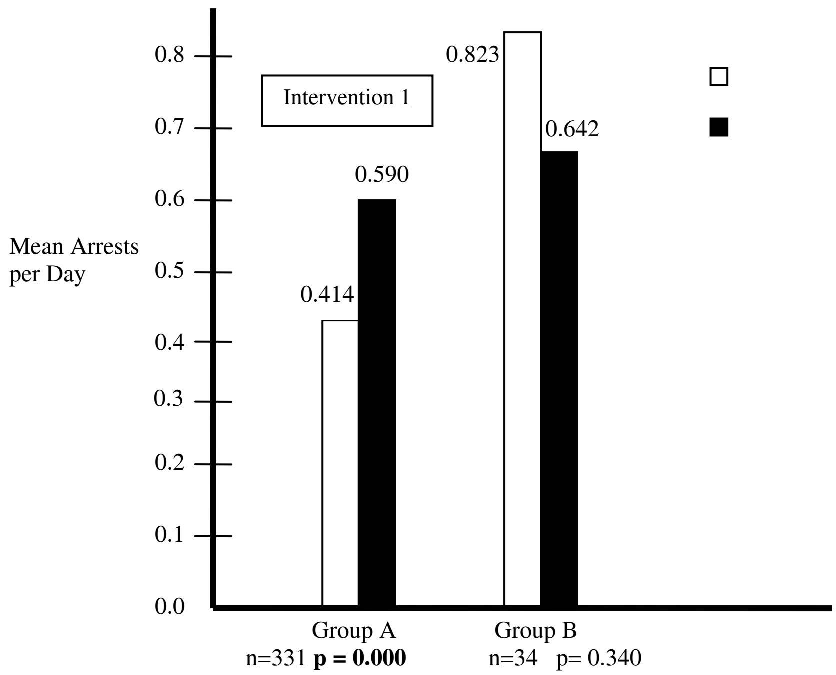

Figure 16: Pre-Post Mean Arrests/Day Intervention I

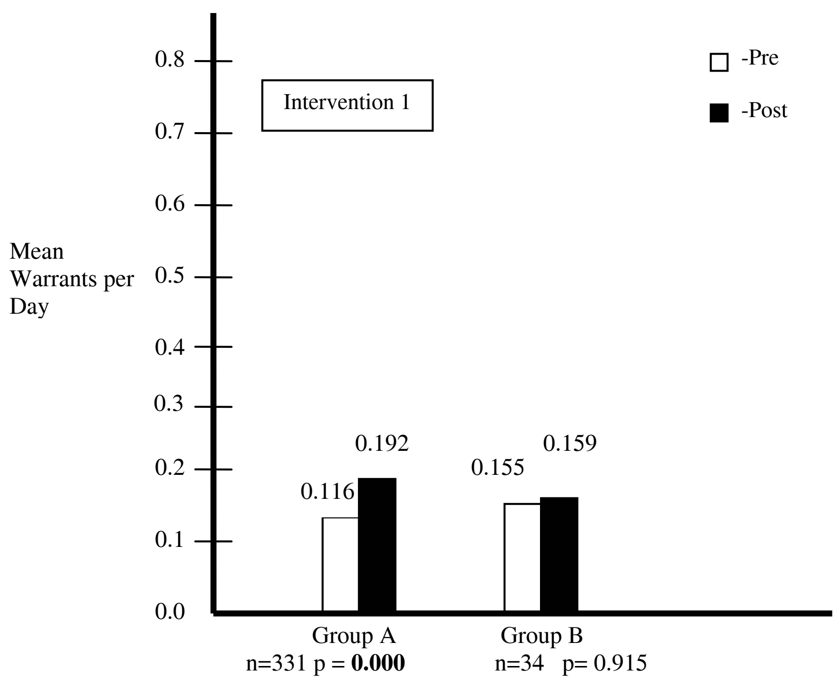

Figure 17: Pre-Post Mean Warrants Served/Day Intervention I

Hypothesis$_{2A}$ shows *no support* for feedback, by the mean number of queries per day, affecting high-risk decision performance from results with Group B in Intervention I (p=0.915). This result is opposite of what was expected however because the significant increase in performance was with the group without computers, Group A (**p=0.000**). Those getting computers, Group B, showed no statistically significant gains in performance outcome measures even though they experienced a statistically significant increase in the mean number of queries per day (feedback).

Hypothesis$_{2B}$ states that *multiple cues, by the number of types of queries, affect high-risk decision performance.*

It is expected that the number of types of queries from different databases allows more clarification of anomalies in high-risk decision-making. Hypothesis$_{2B}$ is not tested in Intervention I where multiple cues affect high-risk decision-making performance because data on types of queries in unavailable.

Hypothesis$_{2C}$ is not tested in Intervention I where an interactive effect of feedback and multiple cues affect high-risk decision-making performance because data on types of queries in unavailable.

<u>Hypothesis$_3$ testing</u>

Hypothesis$_3$ states that *information thresholds moderate performance in high-risk decision-making.* The relationship between the independent variables and performance is non-linear.

There were statistically significant increases in the estimated queries/day for Group B in Intervention I (from secondary analysis) but it was the non-targeted Group A that showed statistically significant performance gains in both mean arrests/day and mean warrants served/day. There were *no statistically significant performance gains* in the group achieving this increase in queries.

Summary of results from Intervention I

Hypothesis$_{1A}$ is supported in Intervention I that information latency affects feedback in high-risk decision-making. Very large changes in latency in Group B resulted in statistically significant (p=**0.000**) changes in the mean number of queries per day.

Hypothesis$_{1B}$ is not tested in Intervention I that information latency affects multiple cues in high-risk decision-making because data on types of queries in unavailable.

Hypothesis$_{2A}$ shows *no support* for feedback, by the mean number of queries per day, affecting high-risk decision performance from results with Group B in Intervention I. This result is opposite of what was expected however because the significant increase in performance was with the group without computers, Group A. Those getting computers, Group B, showed no statistically significant gains in performance outcome measures.

Hypothesis$_{2B}$ is not tested in Intervention I that multiple cues affect high-risk decision-making performance because data on types of queries in unavailable.

Hypothesis$_{2C}$ is not tested in Intervention I where an interactive effect of feedback and multiple cues affect high-risk decision-making performance because data on types of queries in unavailable.

Hypothesis$_{3}$ is not supported in Intervention I because the group (B) achieving the significant increase in feedback, mean numbers of queries per day, showed no corresponding change in performance. A statistically significant change in performance was expected for the group getting computers, since they had more than tripled their mean queries per day. This was not the case.

Intervention II – June – August 1998

Additional wireless computers with similar communication and performance specifications to those in Intervention I were issued from June through August 1998 to a larger group of officers. User query records reported 34 uniformed patrol officers still performing wireless computer queries with the computers from Intervention I (Group B), 192 officers using the Intervention II computers (Group C, a subset of A), and 166 without computers (Group D; another subset of Group A) (n= 392). Only 129 wireless laptops were issued during Intervention II but records indicate 63 additional users who would borrow the computers when the original user is off duty (*i.e.* other duty hours, on vacation, etc.).

Table 10: Intervention II Time Periods and Counts (% of sample)

Intervention	**Pre-Test Period**	**Treatment Period**	**Post-Test Period**	**(%) Of N**	**Description**
II (Group B)	10/1/97 to 12/31/97	June – August 1998	10/1/98 to 12/31/98	392 (97%)	34 from I
(Group C)	”	”	”	”	192 new users
(Group D)	”	”	”	”	166 without

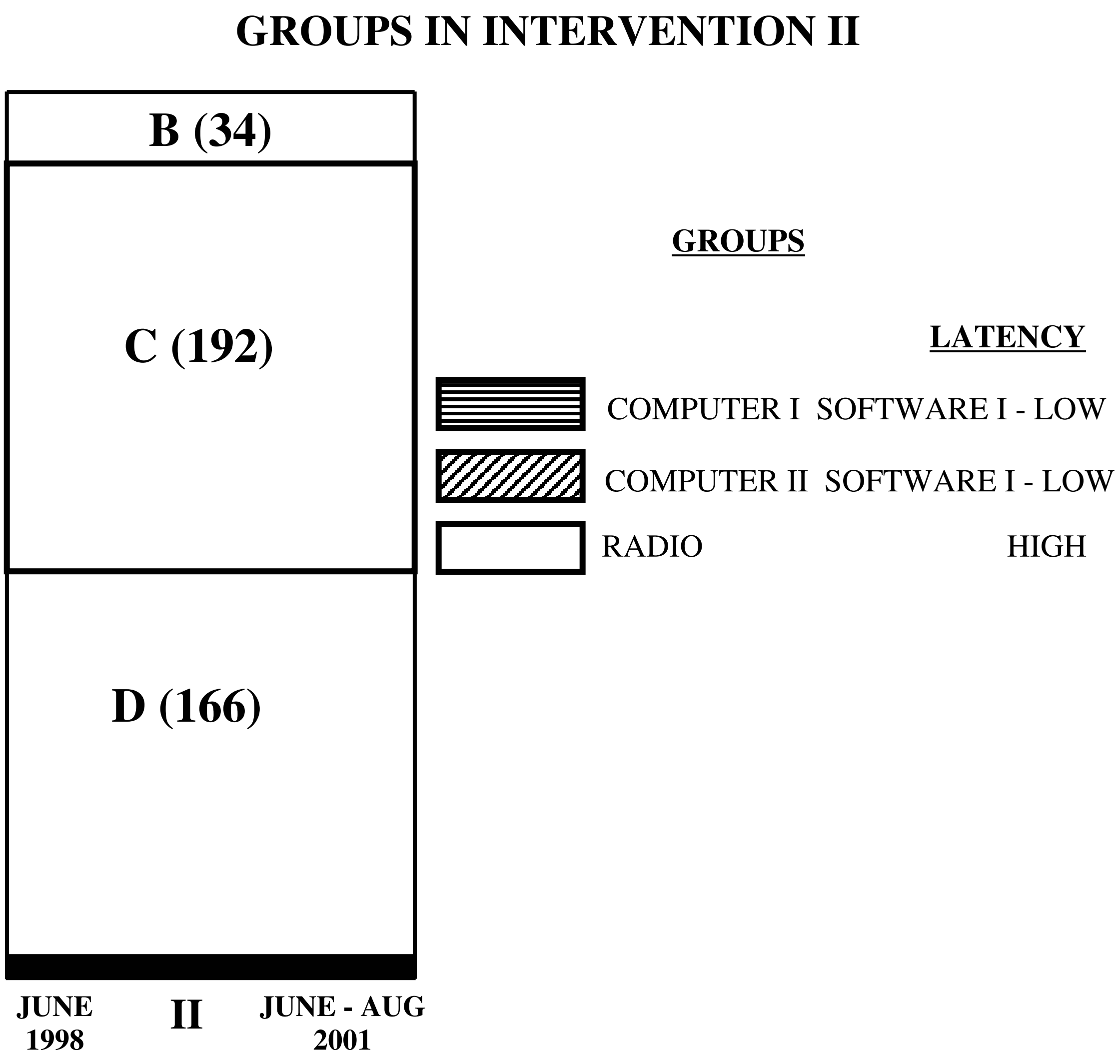

Figure 18: Intervention II Groups

Table 11: Overview of Information Latency changes for Intervention II

Group Identification	**Pre-treatment**	**Post-treatment**	**Δ Comparison of means (p)**
II (Group B)	Wireless queries	Wireless queries	↔1.000
(Group C)	Est. Radio with queuing	Wireless queries	**↓0.000**
(Group C)	Est. Radio w.o. queuing	Wireless queries	**↓0.000**
(Group D)	Est. Radio with queuing	Est. Radio with queuing	↔1.000

Table 11 (above) demonstrates that there was no change in latency for Groups B (those already with computers) and D (those without a computer) but there was statistically significant change in drop in latency for Group C, the group getting new computers in Intervention II.

Hypothesis$_{1A}$ testing for Intervention II

Hypothesis1$_A$ states that information latency affects feedback in high-risk decision-making. The reduction in query latency should increase the number of queries for the officers across the interventions regardless of query type. Pre- and post- data by mean queries per day is not readily available for II. There were records for queries post Intervention II as will be described in a secondary analysis that follows.

Secondary analysis of mean number of queries with Intervention II Groups B

While pre- and post mean queries/day data were not available until post- Intervention II, the record of wireless queries was available showing the number of queries generated by each of the 34 officers in Intervention I (Group B). Their mean queries per day are then estimated to be (79,008 queries recorded for these officers/ (34 officers * 121 days worked)) equal to 19.2 queries per officer per day. The mean queries per day were available for the period October to December 1999 and were equal to 11.14 mean queries per day ($\sigma = 42.406$) for the 34 officers in Group B. Using a paired t-test from the 19.2 pre-Intervention II estimate and comparing to the mean of 11.14 from the 1999 data shows no statistically significant change in numbers of queries for Group B in Intervention II ($p=0.282$). There was no statistically significant change in latency for Group B and no statistically significant change in the number of queries per day.

Secondary analysis of mean number of queries with Intervention II Groups C

While pre- and post mean queries/day data were not available until post- Intervention II, a sampled estimate of radio queries per officers without computers is available showing the number of queries per officer in 1998 is six queries per officer (n=30, Metro Police Planning and Research Division). The mean queries per day were available for the period October to December 1999 and 2000[9]. These were equal to 12.67 mean queries per day ($\sigma = 21.7$) for the 192 officers in Group C in 1999 and 11.88 mean queries per day ($\sigma = 19.03$) for the 192 officers

[9] Note that this data is one and two years after the introduction of the computers for these officers

in Group C during 2000. Paired t-tests from the sample estimate of six queries per day pre-Intervention II and comparing to the means in the 1999 and 2000 data shows statistically significant increases in numbers of queries for Group C in Intervention II.

Hypothesis$_{1A}$ is supported in Intervention II that information latency affects feedback in high-risk decision-making. Statistically significant changes in latency in Group C resulted in statistically significant increases in the mean number of queries per day (p=**0.000**).

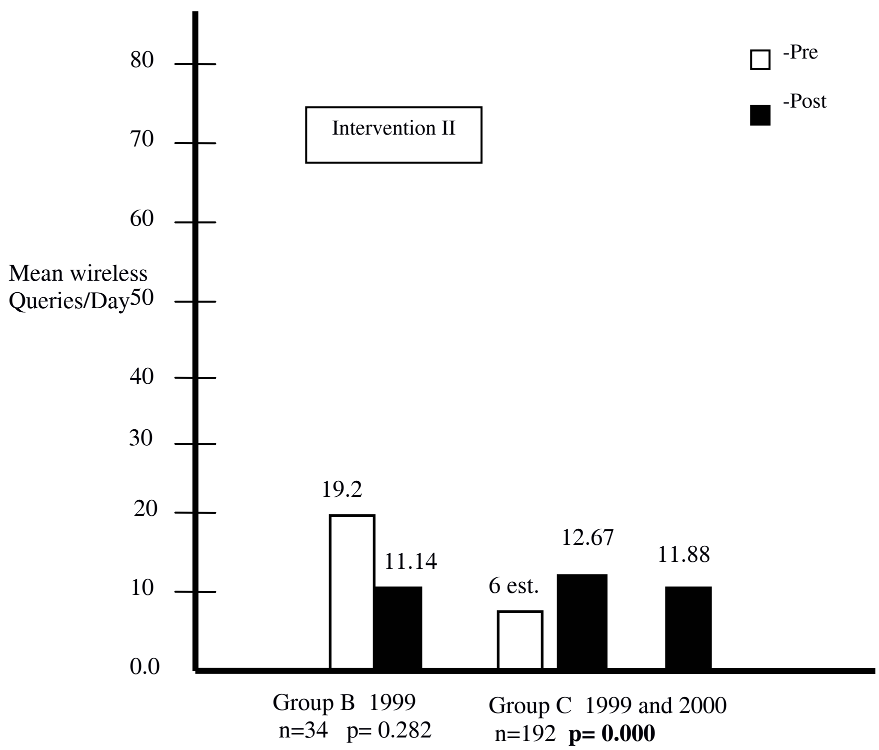

Figure 19: Mean Queries per day for Intervention II

<u>Hypothesis1_B testing</u>

Hypothesis1_b states that *information latency affects multiple cues in high-risk decision-making*. The change in query latency is expected to change the number of types of queries across the interventions. Paired t-tests could not be performed on Intervention II by mean type of queries because mean types of queries data was not available until post- Intervention II.

Hypothesis$_{2A}$ testing in Intervention II

Hypothesis$_{2A}$ states that *feedback*, measured *by the mean number of queries per day, affects high-risk decision performance.*

It is expected an increase in the mean number of queries per day allows more clarification of anomalies in high-risk decision-making. Outcome performance measures within groups, pre and post intervention, demonstrated no statistically significant change in Group B (no change in latency) in either arrests per day or warrants served per day (p=0.999 and p=0.676 respectively) using paired t-tests. T-test comparisons across the groups in Intervention II show a statistically significant increase in arrests per day for the 192, Group C officers using new computers (radio to fast wireless), Group C (p = **0.004**) with a mean arrest rate = 0.513 arrests/day (pre) and post = 0.671 arrests/day though there is no corresponding statistically significant increase in pre to post warrant service rates (p=0.244). Paired t-tests demonstrated no statistically significant change in Group D (radio queries to radio queries) in either arrests per day or warrants served per day (p=0.169 and p=0.389 respectively - see Figure 20).

With Group B, there is no change in latency and there is no change in performance. With Group C, there is a statistically significant decrease in latency and a corresponding increase in arrests as expected. The failure to achieve an increase in the mean number of warrants served with an increase in the numbers of queries is not expected. The Group D had no change in performance as expected since there was no apparent change in latency during this intervention.

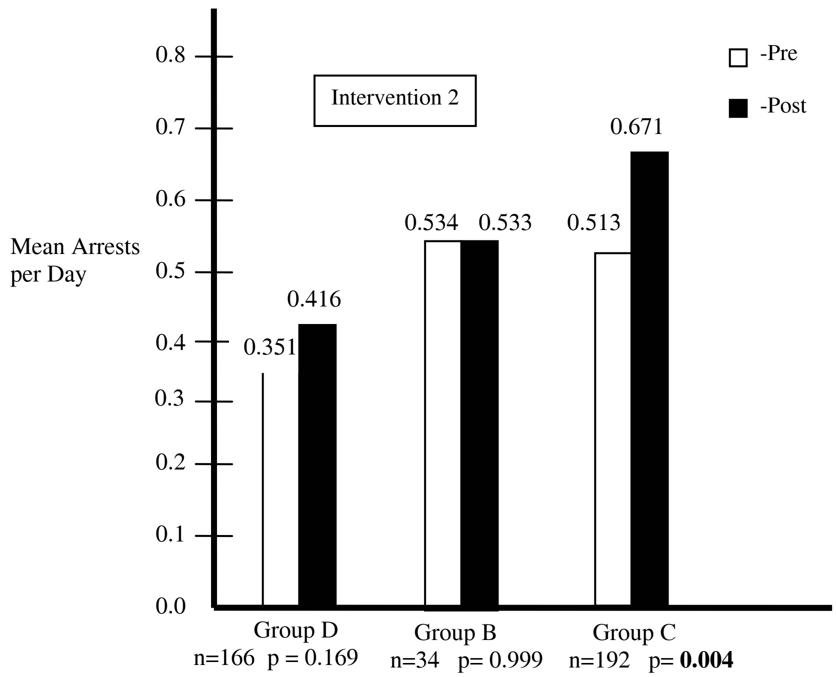

Figure 20: Pre-Post Mean Arrests/Day Intervention II

Paired t-tests across Intervention II showed no statistically significant increases in mean warrant service rates within any group (see Figure 21 below).

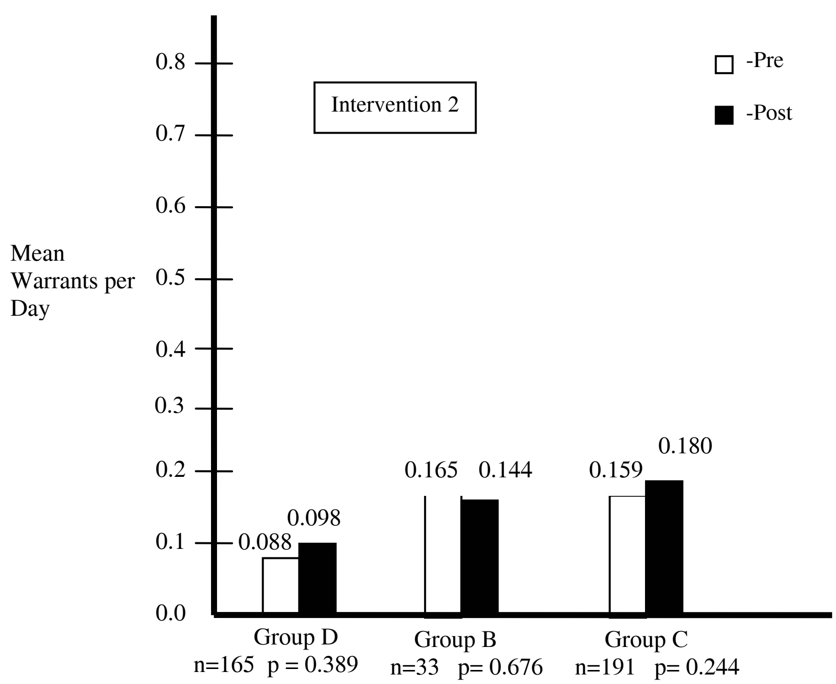

Figure 21: Pre-Post Mean Warrants Served/Day Intervention II

Hypothesis$_{2A}$ shows *support* for feedback, by the mean number of queries per day, affecting high-risk decision performance from results with Group C in Intervention II.

This result only applies to mean arrests per day and not mean warrants served per day. Those already with computers, Group B, showed no statistically significant gains in performance outcome measures and they experienced no statistically significant increase in the mean number of queries per day (feedback).

Hypothesis$_{2B}$ states that *multiple cues, by the number of types of queries, affect high-risk decision performance.*

Hypothesis$_{2B}$ and $_{2C}$ are not tested in Intervention II because data on types of queries in unavailable.

Hypothesis$_{3}$ testing

Hypothesis$_{3}$ states that *information thresholds moderate performance in high-risk decision-making.*

There were statistically significant increases in queries/day for Group C in Intervention II (from secondary analysis) and Group C showed statistically significant performance gains in mean arrests/day but not with mean warrants served/day. This supports significant changes in information resulting in changes in high-risk decision-making performance outcome measures. Hypothesis$_{3}$ is supported for Group C in Intervention II.

Summary of results from analysis of hypotheses for Intervention II

Hypothesis$_{1A}$ is supported in Intervention II that information latency affects feedback in high-risk decision-making performance. Statistically significant changes in latency in Group C resulted in statistically significant (p=0.000) changes in the mean number of queries per day.

Hypothesis$_{1B}$ is not tested in Intervention II that information latency affects multiple cues in high-risk decision-making because data on types of queries in unavailable.

Hypothesis$_{2A}$ is *supported* for feedback, by the mean number of queries per day, affecting high-risk decision performance from results with Group C in Intervention II.

Group C showed statistically significant gains in the number of queries and showed statistically significant increases in their mean number of arrests per day. Group C did not show an increase in their mean number of warrants served per day.

Hypothesis$_{2B}$ is not tested in Intervention II that multiple cues affect high-risk decision-making performance because data on types of queries in unavailable.

Hypothesis$_{2C}$ is not tested in Intervention II where an interactive effect of feedback and multiple cues affect high-risk decision-making performance because data on types of queries in unavailable.

Hypothesis$_3$ is supported in Intervention II because Group C achieved significant increases in feedback, by mean numbers of queries per day, and showed a corresponding improvement in performance outcome measures.

Intervention III – June – August 2001

Five hundred (500) additional computers were introduced over a six month period starting in June 2001. This intervention is recorded for June through August 2001. During this same period, computers issued in 1997 were retired and replaced while computers issued in 1998 were either upgraded with new software or retired and replaced with 2001 computers. There were 124 officers getting issued computers for the first time (Group E, a subset of D), 138 officers who use Intervention II computers and are getting replacements (Group F, a subset of C), 30 officers upgrading their Intervention II computers (Group G, a subset of C) with software, and 12 officers replacing their Intervention I computer with new computers (Group H, a subset of B) (n=304). Computers continued to be issued to all officers in patrol beyond the date of this research.

GROUPS IN INTERVENTION III

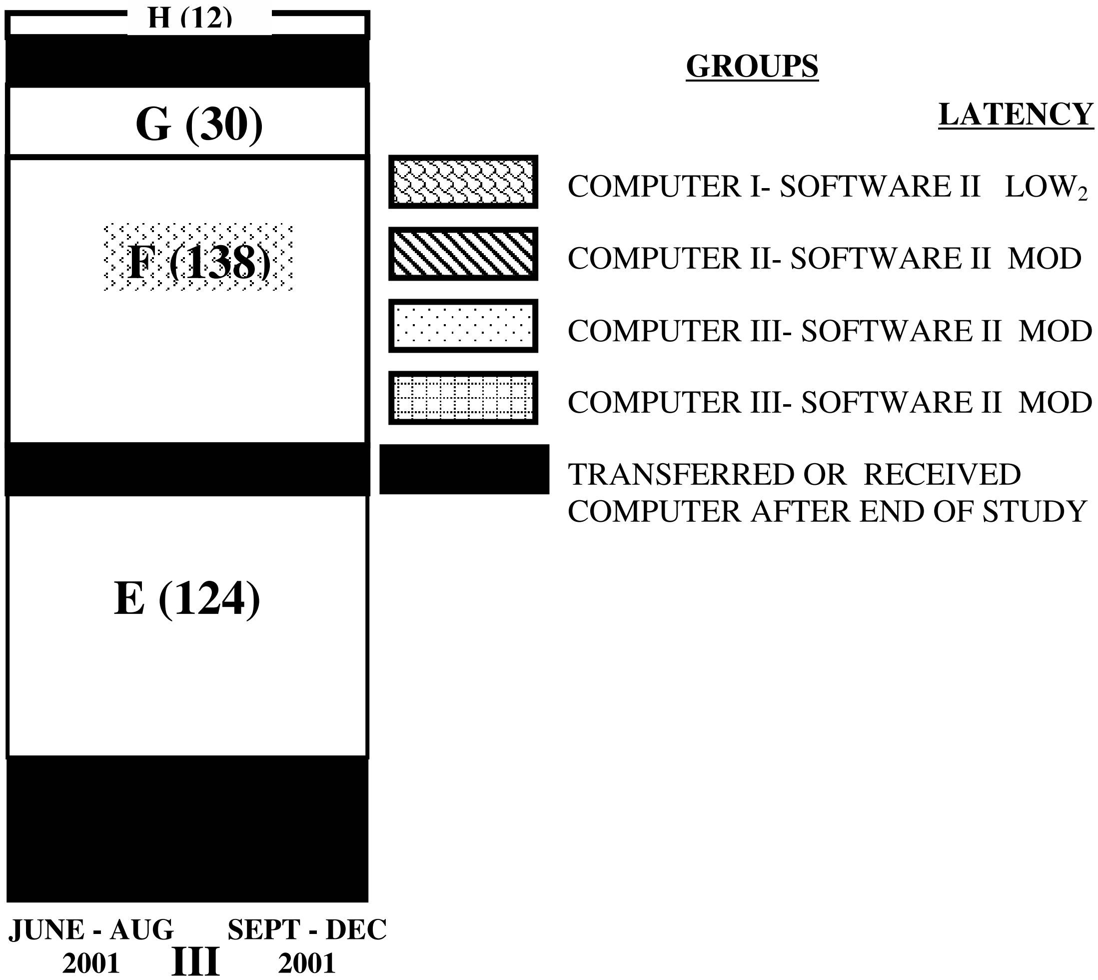

Figure 22: Intervention III Groups

Table 12: Group III Time Periods and Counts (% of sample)

Intervention	**Pre-Test Period**	**Treatment Period**	**Post-Test Period**	**(%) of N**	**Description**
III (Group E)	10/1/00 to 12/31/00	June – August 2001	10/1/01 to 12/31/01	304 (75%)	124 new users
(Group F)	”	”	”		138 replacing II
(Group G)	”	”	”		30 upgrading II
(Group H)	”	”	”		12 replacing I

Hypothesis$_{1A}$ testing for Intervention III

Hypothesis1$_A$ states that information latency affects feedback in high-risk decision-making. The reduction in query latency should increase the number of queries for the officers across the interventions regardless of query type. Paired t-tests were run on Intervention III. Query data was available for Groups F, G, and H but only post intervention for Group E. Groups F and G will be compared but Group H (12 officers replacing Intervention I computers) was not included because of the low sample size.

Table 13: Radio and computer surveyed mean query times

Type Query	Mean Radio Latency 5 Min Queuing (1998) + Radio Query = Total (Very High) Latency				Mean Latency Int. I, II (1998) (Low)	Mean Latency Intervention III A (2001) (Moderate)
	Queue	**Radio**	**Total**	**n**		
Vehicle Registration	5:00	1:58	6:58	94	0:02 (n=30)	0:5 (n=48)
Criminal History	5:00	3:50	8:50	188	0:04 (n=30)	0:20 (n=48)
Driver's License (State)	5:00	2:56	7:56	57	0:12 (n=30)	0:25 (n=48)
Warrant Check	5:00	2:30	7:30	175	0:04 (n=30)	0:20 (n=48)

Table 14: Latency Changes for Intervention III

Group Identification	**Pre-treatment**	**Post-treatment**	**Δ Comparison of means (p)**
III (Group E)	Est. Radio with queuing	Slower wireless queries	**↓0.000**
(Group F)	Wireless queries	Slower wireless queries	**↑0.0012**
(Group G)	Wireless queries	Slightly slower wireless	Not available
(Group H)	Wireless queries	Slower wireless queries	**↑0.0012**

Latency changes *were statistically significant* for Groups E and F. Group G latency changes were unavailable and Group H is of insufficient sample size.

Data demonstrates that moderate and slight changes in information latency showed no statistically significant change in mean queries per day across the intervention with p=0.399 for Group F (137 officers replacing Intervention II computers) nor for Group G (see Figure 23).

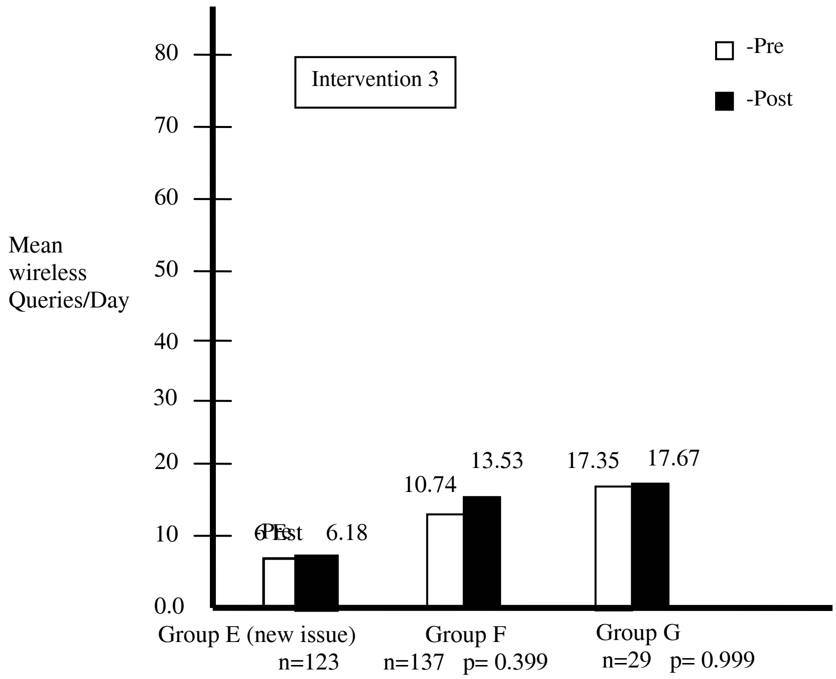

Figure 23: Mean Queries per Day for Intervention III

Pre-Intervention mean radio queries/day data is not available for Group E. This group experienced a large reduction in latency (going from very high latency radio queries to moderate latency wireless queries). Post Intervention III, Group E revealed a mean = 6.18 wireless queries/day. A secondary analysis is necessary since the pre-Intervention III level of wireless queries/day for Groups F and G are at high level of queries per day pre-intervention (10.74 and 17.35 respectively) and pre-intervention number of queries was not available and had to be estimated. Group H was not included due to low sample size (n=12).

The *estimated* mean radio query/day rate/officer is, approximately, six (6) queries/day; a rate from the data collected by the police planning and research section. Six queries per day pre-computer were estimated in 1997, 1998 and 5.77 queries per officer per day in 2001. The 2001 estimate is based on a sample of 48 officers performing 277 queries (mean = 5.77) Using a comparison of means in a large-samples ($n_1 \geq 20$, $n_2 \geq 20$) significance test (Agresti and Finlay 1997) it is possible to calculate the significance of changing from an estimated mean of 5.77 radio queries/day to a mean of 6.18 wireless queries per day. The standard deviation for the 2001 sample is 5.635 but conservatively, even using the largest standard deviation of any of the groups, would not make the differences, in means, statistically significant. The standard deviation post-Intervention III for Group E is higher than the standard deviation for any group *pre-* Intervention III. The value of 5.6 for the *sampled* standard deviation of mean radio queries/day is used for illustration.

$$\hat{\sigma}\bar{Y}2 - \bar{Y}1 = \sqrt{\frac{s1^2}{n1} + \frac{s2^2}{n2}}$$

$$\hat{\sigma}\bar{Y}2 - \bar{Y}1 = \sqrt{\frac{5.635^2}{123} + \frac{19.5^2}{123}}$$

$$\hat{\sigma}\bar{Y}2 - \bar{Y}1 = 3.35$$

$$z = \frac{\bar{Y}2 - \bar{Y}1}{\sqrt{\frac{s1^2}{n1} + \frac{s2^2}{n2}}}$$

$$z = \frac{6.15 - 5.77}{3.35} = 0.113$$

$$p \cong 0.456$$

There was no statistically significant change in the mean number of queries with large changes in latency for Group E in Intervention III. This is the opposite of what is expected.

Hypothesis$_{1A}$ is not supported in Intervention III that information latency affects feedback in high-risk decision-making. Large, moderate and slight changes in latency resulted in *no statistically significant changes* in the mean numbers of queries per day in Intervention III.

Hypothesis$_{1B}$ testing

Hypothesis$_{1B}$ states that *information latency affects multiple cues in high-risk decision-making.* The change in query latency is expected to change the number of types of queries across the interventions.

Pre-intervention *radio* queries for Group E were used in a secondary analysis of means. Groups F and G in Intervention III were tested. Group H, with only 12 officers was deemed to small a sample. Paired t-tests pre- and post- for Group F by mean types of queries demonstrated a statistically significant change (p=**0.005**) across the intervention with a moderate increase in latency. Group G demonstrated no statistically significant change in mean types of queries across the intervention (p=0.617) (see Figure 24).

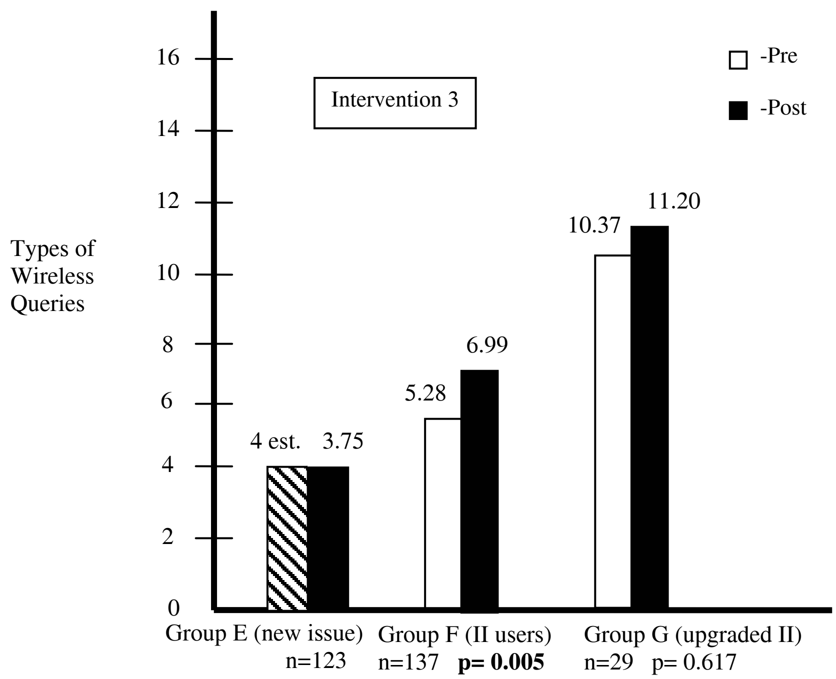

Figure 24: Types of Queries Intervention III

Secondary analysis of number of types of queries with Intervention III Group E

Pre-Intervention mean types of queries data is not available for Group E. This group experienced a large reduction in latency (going from very high latency radio queries to moderate latency wireless queries). Post Intervention III, Group E revealed a mean = 3.75 types of queries. A secondary analysis is necessary since the pre-Intervention III types of queries for Groups F and G are at high level of mean types of queries pre-intervention (5.28 and 10.37 respectively). The estimated *mean radio types of queries* /officer = four (4) types of queries. Using a comparison of means in a large-samples ($n_1 \geq 20$, $n_2 \geq 20$) significance test (Agresti and Finlay 1997) it is possible to calculate the significance of changing from an estimated mean of 4 radio types of queries to mean of 3.75 wireless types of queries with Group E. The standard deviation for the 1998 estimate of 4 types of queries is not available but conservatively, using the largest standard deviation of any of the groups, would not make the differences, in means, statistically significant. The standard deviation post-Intervention for Group E is higher than the standard deviation for any group pre- Intervention. The largest, most conservative pre-Intervention III value of 8.2 for the *estimated* standard deviation of mean types of queries is used for illustration. Data shows the standard deviation for mean types of queries post- Intervention III for Group *E* to be 4.986 (n=123).

$$\hat{\sigma}\overline{Y}_2 - \overline{Y}_1 = \sqrt{\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}}$$

$$\hat{\sigma}\overline{Y}_2 - \overline{Y}_1 = \sqrt{\frac{3.75^{\,2}}{123} + \frac{8.2^{\,2}}{123}}$$

$$\hat{\sigma}\overline{Y}_2 - \overline{Y}_1 = \sqrt{.202 + .54}$$

$$\hat{\sigma}\overline{Y}_2 - \overline{Y}_1 = \sqrt{0.749}$$

$$\hat{\sigma}\overline{Y}_2 - \overline{Y}_1 = 1.067$$

$$z = \frac{\overline{Y}_2 - \overline{Y}_1}{\sqrt{\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}}}$$

$$z = \frac{3.75 - 4}{1.067} = -0.234$$

$$p \cong 0.409$$

Hypothesis$_{1B}$ is *not supported* that information latency affects multiple cues in high-risk decision-making for Group E. This is the opposite of what was expected.

Hypothesis$_{1B}$ shows *support* that information latency affects multiple cues in high-risk decision-making for Group F.

Hypothesis$_{2A}$ testing

Hypothesis$_{2A}$ states that *feedback*, measured *by the mean number of queries per day, affects high-risk decision performance.*

It is expected an increase in the mean number of queries per day allows more clarification of anomalies in high-risk decision-making. Data from the three interventions should demonstrate a change in the performance outcome measures across each intervention for the computers users with changes in mean number of queries per day.

Intervention III outcome comparison within groups

Paired t-tests compare the mean arrests/day, and mean warrants/day of the three groups across the pre and post Intervention III period versus changes in numbers of queries. Group E (new issue computers) showed no statistically significant changes in mean arrest rate/day ($p=0.085$) with no statistically significant change in the mean queries/day ($p=0.274$). Group F (1998 computer users getting new replacements) showed a significant *drop* in their mean arrest rate per day going from 0.506 arrests per day to 0.222 arrests per day post intervention ($p=$**0.004**) with no statistically significant change in mean queries/day ($p=0.399$). Group G experienced a slight increase in latency, showed no statistically significant change in mean arrests per day ($p=0.357$), and showed no statistically significant change in mean queries per day ($p=0.999$). Group H was not included because of small ($n=12$) sample size (see figure 25).

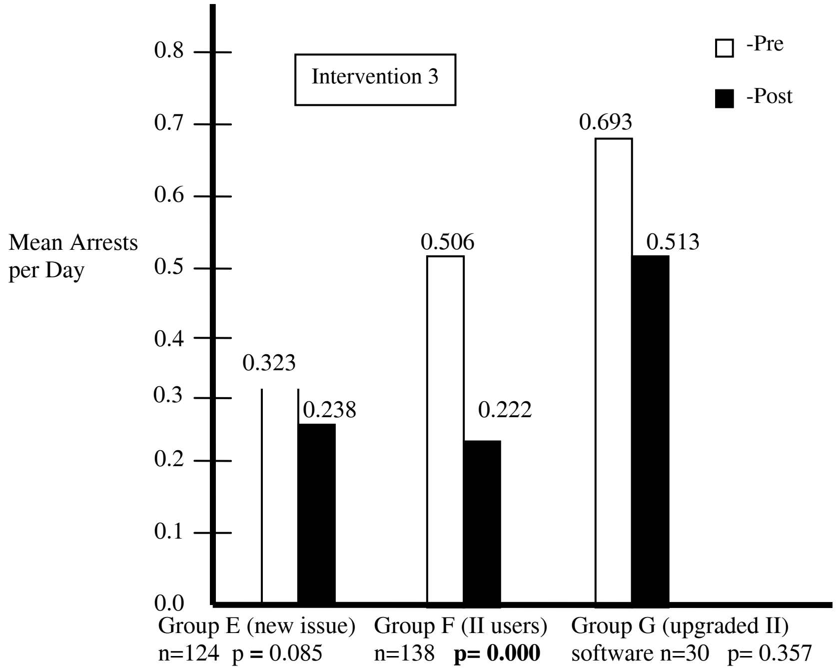

Figure 25: Pre-Post Mean Arrests/Day Intervention III

There were also no statistically significant changes by mean warrants served/day across Intervention III (Figure 26) for Group E (p=0.345) with no statistically significant change in the mean queries/day (p=0.274), for Group F (p=0.369) with no statistically significant change in mean queries/day (p=0.399), or for Group G (p=0.927) and showed no statistically significant change in mean queries per day (p=0.999). Group H was not included due to inadequate sample size (see figure 22 for Intervention III groups).

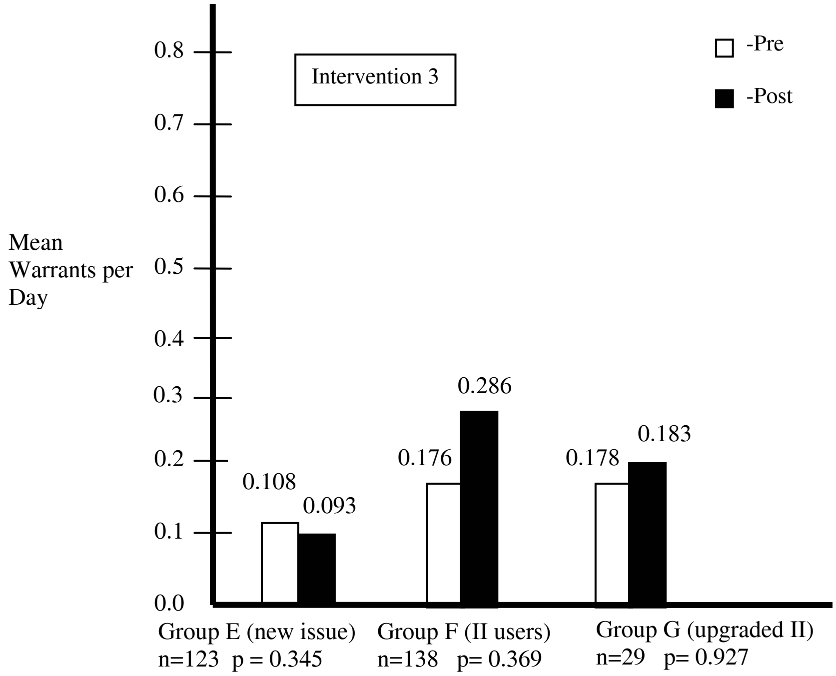

Figure 26: Pre-Post Mean Warrants Served/Day Intervention III

Hypothesis$_{2A}$ shows *no support* for feedback, by the mean number of queries per day, affecting high-risk decision performance from results with Groups E, F, and G in Intervention III.

Hypothesis$_{2B}$ testing

Hypothesis2B states that *multiple cues, by the number of types of queries, affect high-risk decision performance.*

It is expected that the number of types of queries from different databases allows more clarification of anomalies in high-risk decision-making.

Group E (new issue computers) showed no statistically significant changes in mean arrest rate/day (p=0.085) with no statistically significant change in the mean types of queries (p=0.409) through secondary analysis by comparisons of means. Group F showed a significant drop in their mean arrests/day (p=**0.004**) with a statistically significant change in mean types of queries (p=**0.005**). This will be discussed later. Group G experienced a slight increase in latency, showed no statistically significant change in mean arrests per day (p=0.357) and showed no statistically significant change in mean types of queries (p=0.617).

There were also no statistically significant changes by mean warrants served/day across Intervention III for Group E (p=0.345) with no statistically significant change in the mean types of queries (p=0.409), for Group F (p=0.369) with a statistically significant change in mean types of queries (p=**0.005**), and for Group G (p=0.927) and showed no statistically significant change in mean types of queries (p=0.617) (refer to figure 20).

Hypothesis$_{2B}$ is *not supported* for multiple cues, number of types of queries, affecting high-risk decision performance with Intervention III Group E and G.

Intervention Groups E and G demonstrated no statistically significant change in types of queries and no statistically significant changes in mean arrests/day. Group F demonstrated a statistically significant increase in the number of types of queries (p=**0.005**) but with no statistically significant effect in mean warrants served/day (p=0.369).

Hypothesis$_{2B}$ is *supported* for multiple cues, number of types of queries, affecting high-risk decision performance with Intervention III Group F.

This group demonstrated a statistically significant increase in the number of types of queries (p=**0.005**) but the effect appeared in the *opposite* direction that was expected with a statistically significant drop in mean arrests/day (p=**0.000**).

Hypothesis$_{2C}$ testing

Hypothesis$_{2C}$ states that *there is an interactive effect between feedback and multiple cues that affect high-risk decision performance.*

It is expected that the number of queries and types of queries from different databases allows more clarification of anomalies in high-risk decision-making.

There is a statistically significant increase in types of queries over Intervention III by Group F resulting in a statistically significant *decrease* in mean arrests/day and no statistically significant change in mean warrants served/day. This is opposite the expected effect. Data did not show any statistically significant changes in mean types of queries except with group F in Intervention III. There was no data available to support any combined statistically significant changes by mean queries/day and mean types of queries during the same intervention.

Hypothesis$_{2C}$ *is supported* for an interactive effect between feedback and multiple cues that affect decision performance in high-risk decision-making with Group F.

Hypothesis$_3$

Hypothesis$_3$ states that *information thresholds moderate performance in high-risk decision-making.* The relationship between the independent variables and performance is non-linear.

Slight increases in information latency (Intervention III Group G), moderate increases in information latency (Group F) and high decreases in information latency (Intervention III Group E) resulted in no statistically significant changes in mean queries per day.

Moderate increases in latency resulted in statistically significant increases in the mean types of queries in Intervention III, Group F which led to statistically significant decreases in mean arrests/day.

Any effects appear *non-linear between independent variables*.

Hypothesis$_3$ is *supported* for information thresholds moderating performance in high-risk decision-making.

A summary of the hypothesis testing within groups

The results of testing do not give a clear understanding of the effects of variables. Results from one intervention clash with other intervention results. A summary of hypothesis results follows (Table 15).

Table 15: Summary of Hypotheses Testing In High-Risk Decision Making

Hypothesis	Dependent variables	Results	Which Group(s)
1a	information latency affects feedback	not supported	Groups E, F, G
1a	information latency affects feedback	supported	Group B, C
1b	information latency affects multiple cues	not supported	Group E
1b	information latency affects multiple cues	supported	Group F
1b	information latency affects multiple cues	not supported	Group G
2a	feedback affects decision performance	not supported	Groups E, F, G
2a	feedback affects decision performance	supported	Group A, C
2b	multiple cues affect decision performance	not supported	Groups E, G
2b	multiple cues affect decision performance	supported	Group F
2c	Interactive effect of feedback and multiple cues affecting decision performance	supported	Groups C, F
3	information thresholds moderating performance	supported	Group C, F

A pattern emerges when the results above are added to the model (Figure 27). It is clear that the group makeup plays an important role in performance outcome measures.

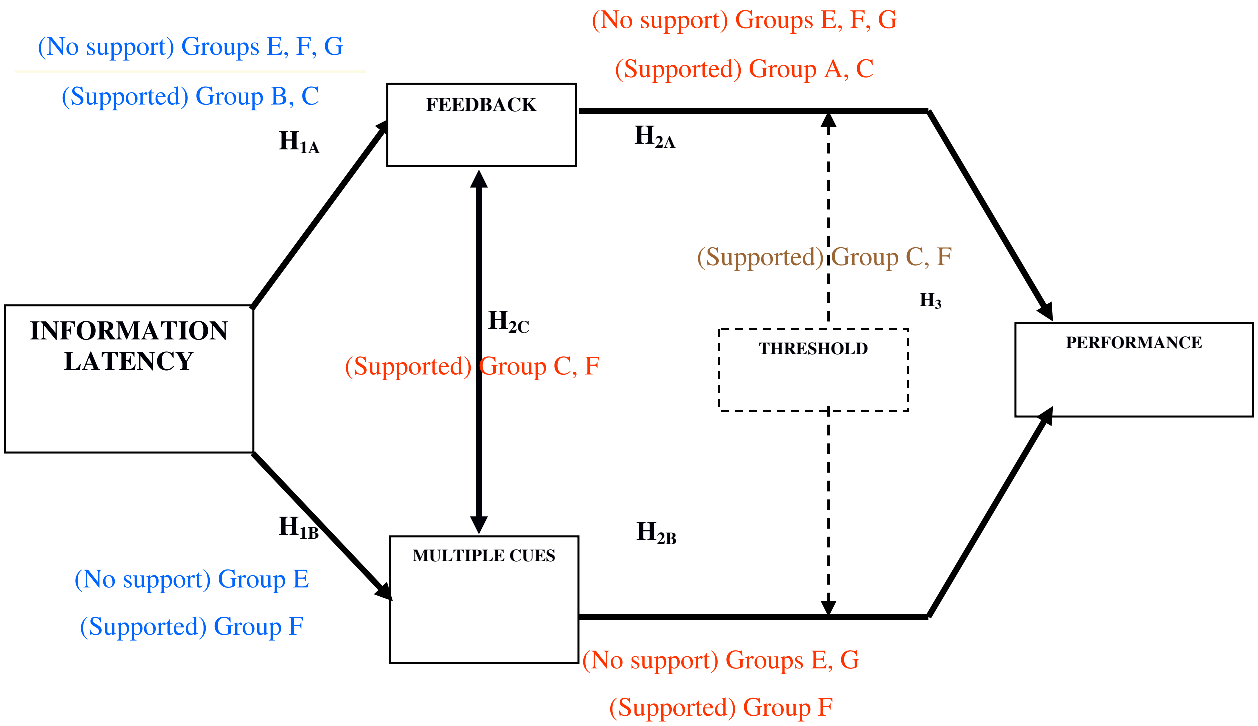

Figure 27: Summary of results of within group hypotheses testing

Information latency does not appear to be the driving force for feedback as the results were mixed. Feedback and multiple cues do not appear be the driving force for performance as those results were also mixed. There appears to be some form of interactive effect between feedback and multiple cues but it is not clear since the mean number of types of queries increased when the latency increased (opposite of expected effect). The same is true for threshold effects since even though there significant increase in latency for Group F in Intervention III, they demonstrated a significant increase in types of queries used yet a decrease in arrests. Group E experienced a significant decrease in latency yet showed no performance changes whatsoever.

Statistical analysis, through paired t-tests, is not sufficient to explain group effects across the interventions. ANOVA comparison among groups may better explain this non-linearity. The group similarities and differences may play an important role in performance outcome measures and ANOVA comparisons demonstrate these performance characteristics among groups. Extended analysis of groups is necessary to better explain these results and is discussed in the next chapter.

CHAPTER V

EXTENSION OF RESEARCH

The results suggest that group characteristics are more important to explain performance outcomes measures than information latency. Information latency does not appear to be the driving factor. It must be remembered that these groups were not randomly assigned computers. This chapter will examine those group characteristics. This chapter will attempt to analyze these group characteristics.

The results for all interventionsI were mixed at best. Expected changes appeared but in the opposite direction (Group A, B, and F). Changes did not occur when expected (Group E). There was also an anomaly in Intervention I that was examined next. Significant drops in latency resulted in performance gains in the non-computer group (Group A). No other group without computers showed any performance changes during this study.

Groups A and C (and F) are evaluated for performance traits

Group A showed statistically significant gains in performance outcome measures when they had no computer assigned to them. It is likely that some process change occurred when the first computers were assigned in 1997 and 1998.

Intervention I data was reexamined to determine if Group C was comprised of the high performers from Group A. Officers who would later get computers in Intervention II were separated from the officers in Group A that would not get computers until after Intervention II. Group C, who were without computers after Intervention I, were compared in a paired t-tests to the remainder of Group A, also without computers. The Group C officers, without computers after intervention I ($n=192$), statistically significantly improved in mean arrests/day and mean warrants served/day ($p=0.000$ and $p=0.001$ respectively) in pre- post- comparison. The remainder of Group A ($n= 139$) did not statistically significantly improve in either arrests/day or mean warrants served/day ($p=0.252$ and $p=0.014$ respectively) with pre- and post- Intervention I paired t-tests.

Among Group Testing

Among group comparisons are performed using one-way analysis of variance. Post hoc interval analysis is performed when statistically significant differences occur among groups.

Intervention I is examined through one way analysis of variance, ANOVA, among the two groups by pre- and post- log mean arrests/day and log mean warrants served/day.

Intervention II is also examined by ANOVA using log mean arrests/day and log mean warrants served/day with the three groups in Intervention II with post hoc Bonferroni and Tukey HSD interval analysis to examine specific differences in the groups (see Table 16).

Table 16: Statistical Tests of Hypotheses among Groups

INTERVENTION	MEASURES USED	SPECIFIC TESTS	GROUPS
I	Log Mean arrests/day	ANOVA	A, B
	Log Mean warrants served/day	ANOVA	A, B
II	Log Mean arrests/day	ANOVA & Post Hoc Bonferroni & Tukey HSD	B, C, D
	Log Mean warrants served/day	ANOVA & Post Hoc Bonferroni & Tukey HSD	B, C, D
	Log Mean arrests/day	ANOVA & Post Hoc Bonferroni & Tukey HSD	B, C, D
	Log Mean warrants served/day	ANOVA & Post Hoc Bonferroni & Tukey HSD	B, C, D
III	Log Mean arrests/day	ANOVA & Post Hoc Bonferroni	E, F, G, H
	Log Mean warrants served/day	ANOVA & Post Hoc Bonferroni	E, F, G, H
	Log Mean Queries/day	ANOVA & Post Hoc Bonferroni	E, F, G, H
	Log Mean Types Queries	ANOVA & Post Hoc Bonferroni	E, F, G, H
	Log Mean arrests/day	ANOVA & Post Hoc Tukey B and Scheffe	E, F, G, H
	Log Mean warrants served/day	ANOVA & Post Hoc Tukey B and Scheffe	E, F, G, H
	Log Mean Queries/day	ANOVA & Post Hoc Tukey B and Scheffe	E, F, G, H
	Log Mean Types Queries	ANOVA & Post Hoc Tukey B and Scheffe	E, F, G, H

Intervention III examination also uses ANOVA by log mean arrests/day and log mean warrants served/day but also examines by log mean queries/day and log types of queries. There are four groups in Intervention III and post hoc Bonferroni interval analysis is used to examine differences in the groups when statistically significant differences are found. Post hoc Tukey B and Scheffe analyses are additionally used to examine homogeneous subsets of groups due to the increased numbers of variables in Intervention III.

ANOVA and Post Hoc Comparison among Groups

Two types of evaluation are completed *among* the groups. One-way analysis of variance (ANOVA) is used to examine among group differences in Intervention I, II, and III and post hoc interval analysis is used in Intervention II and III when there are statistically significant differences among groups.

Linear regression and ANOVA analysis was not used with the dependent variables. Paired t-tests, for repeated measures, are equivalent to regression and ANOVA analysis (Agresti and Finlay 1997). The repeated comparison of means is an effort to control potential timing and maturation biases. Also the standard error of $\bar{Y}_2 - \bar{Y}_1$ is smaller using comparison of means than the regression/Pearson correlation. An example with the statistical significance in mean arrests/day changes pre and post intervention II shows $\rho = 0.002$ using a Pearson correlation and $\rho = 0.004$ using an ANOVA comparison of means. Comparison of means is an accurate a method to examine data in this study as linear regression and Pearson correlation.

<u>Group age and experience demographics by intervention</u>

The groups, in all three interventions, showed no statistical significance in either age or experience at the time of that particular intervention (see Table 5 and 6). These are controlled variables to determine if age and experience were major influences on outcomes. Police officers are required to be at least 21 years of age (Gun Control Act of 1968, Title 18, US Code Chapter 44, 922 W 3A; TCA 38-8-101 (1) F) to start and work at least eleven months (Metro Nashville Police Department Field Training Officer, General Order) before they are allowed to patrol by themselves. Police officers typically leave patrol for other departmental assignments after age fifty (50).

One way analysis of variance for the two groups in Intervention I showed no statistical difference in the ages (p=0.839) or experience (p=0.973) of the groups. The 331 without computers in Group A had a mean age of 33.2 years with 7.2 years of experience and the 34 officers using computers in Group B had a mean age of 32.9 years with 7.2 years of experience.

The Intervention II groups are also compared by age and experience. One way analysis of variance showed no statistically significant difference in the ages (p=0.891) or experience (p=0.861) of the groups. The 34 officers in Group B now have a mean age of 33.9 years with 8.2 years of experience and the 192 officers using wireless laptops in Group C have a mean age of 33.4 years with 7.6 years of experience. The 166 without computers had a mean age of 33.7 years with 7.8 years of experience.

The Intervention III groups are then compared in age and experience. One way analysis of variance in ages among the groups is not statistically significant (p=0.466), nor with experience among groups in this intervention (p=0.844). The 124 officers in Group E have a mean age of 36.8 years with 10.9 years of experience. The 138 officers in group F have a mean age of 35.9 years with 10.3 years of experience. The 30 officers in Group G have a mean age of 35.9 years with 10.1 years of experience. The 12 officers in Group H getting their 1997 computers replaced have a mean age of 38.8 years with 12.1 years of experience.

Obviously, age and experience, at a particular period in time can be eliminated as the reason for differences between groups. More in-depth analysis showed experience levels <u>between</u> one intervention and the next, through paired t-tests, demonstrate statistical differences between the level of experience between any intervention (p=0.000). The groups get slightly older and more experienced as they proceed through the years of the study but age and experience do not appear to be the reason for performance changes (see Table 17 and 18).

Table 17: ANOVA Comparison among Groups by Age per Intervention

Intervention	***Mean Age (Yrs)***	***N***	***ρ***
I (Group B)	32.9	34	0.839
(Group A)	33.2	331	
II (Group C)	33.4	192	0.891
(Group B)	33.9	34	
(Group D)	33.7	166	
III (Group E)	36.8	124	0.466
(Group F)	35.9	138	
(Group G)	35.9	30	
(Group H)	38.8	12	

Table 18: ANOVA Comparison among Groups by Experience per Intervention

Intervention	***Mean Experience (Yrs)***	***N***	***ρ***
I (Group A)	7.2	331	0.973
(Group B)	7.2	34	
II (Group B)	8.2	34	0.861
(Group C)	7.6	192	
(Group D)	7.8	166	
III (Group E)	10.9	124	0.844
(Group F)	10.3	138	
(Group G)	10.1	30	
(Group H)	12.1	12	

The groups did not show statistically significant differences in age or experience yet cross comparison of the groups, though analysis of variance (ANOVA), revealed statistically significant differences between the groups through their mean rate of arrests pre (p=**0.000**) and post intervention (p=**0.005**). Comparison of the groups according to their log mean warrant service rate however showed no statistically significant difference pre (p=0.031) or post intervention (p= 0.538) (see Table 19).

Post-Hoc Comparison among Groups: Intervention I

Table 19: ANOVA Comparison of Groups in Intervention I

		Sum of Squares	df	Mean Square	F	Sig.
Log Arrests per day_96	Between Groups	.290	1	.290	17.830	**.000**
	Within Groups	5.815	358	.016		
	Total	6.105	359			
Log Arrests per day_97	Between Groups	.021	1	.021	7.824	**.005**
	Within Groups	.945	358	.003		
	Total	.966	359			
Log warrants per day_96	Between Groups	.016	1	.016	4.671	.031
	Within Groups	1.257	361	.003		
	Total	1.274	362			
Log Warrants per day_97	Between Groups	.002	1	.002	.380	.538
	Within Groups	1.916	362	.005		
	Total	1.918	363			

The Intervention I groups are dissimilar in performance (p=**0.000** and p=**0.005** respectively) as it impacts log mean arrests per day but not statistically significantly dissimilar (p=0.031 and p=0.538 respectively) in log mean warrants served per day pre and post intervention. Group B, getting computers, consists of higher performers, on average, both pre and post Intervention I than Group A, without computers. Both groups are relatively similar performers in warrants served/day whether pre or post intervention.

Post Hoc Comparison among Groups: Intervention II

Examining data from Intervention II (Table 20) it is evident that there are statistically significant differences between the groups in performance. These differences are not due to their age and experience as there is no statistically significant difference between the groups of Intervention II. The groups all showed differences among the outcome measures as shown in this ANOVA table according to their log mean arrests and warrants per day. Though the groups are different in arrests and warrant service outcome measures it does not explain why the hypotheses with Group A and F are supported and others are not (see Table 20).

Table 20: ANOVA Comparison of Groups in Intervention II

		Sum of Squares	Df	Mean Square	F	Sig.
Arrests/day 10-12_1997	Between Groups	.205	2	.103	5.200	**.006**
	Within Groups	7.674	389	.020		
	Total	7.879	391			
Arrests/day 10-12_1998	Between Groups	.403	2	.202	8.264	**.000**
	Within Groups	9.446	387	.024		
	Total	9.849	389			
Warrants per day10-12_1997	Between Groups	.056	2	.028	6.669	**.001**
	Within Groups	1.623	389	.004		
	Total	1.679	391			
Warrants per day10-12_1998	Between Groups	.066	2	.033	6.399	**.002**
	Within Groups	2.007	389	.005		
	Total	2.073	391			

Post Hoc analysis to clarify group differences in Intervention II

A further post hoc Bonferroni interval analysis was necessary to demonstrate differences. It revealed that the greatest difference occurred between the high performing Intervention II Group C and those without computers from the first two interventions (Group D). The 34 officers in Group B (mean= 0.1603), have statistically equivalent pre-intervention mean log arrest/day rates with the 192 officers in Group C (mean=0.1533; p=1.000). The 166 officers in Group D had statistically significant, lower pre-intervention mean log arrest/day rates (mean=0.1082) than the 192 officers in Group C (p=**0.008**). Tukey HSD post hoc analysis of mean log arrests/day post Intervention II not only showed statistically significant differences (p=**0.002**) in performance between the Group D (mean= 0.1236) and the higher performing Group C but showed that the Group C are statistically significantly higher (mean= 0.1811; p=**0.009**) performers than even Group B in post intervention log mean arrests/day. Again, the high performers got the new computers in Intervention II and, since their information latency is greatly reduced, their number of queries should increase. Information latency did not change for the Intervention I computer Group B during Intervention II since the technology remained unchanged.

ANOVA and Post Hoc Comparison among groups in Intervention III

Cross comparison among the groups, though ANOVA, initially revealed statistically significant differences between the groups in all but their log mean rate of warrants served pre intervention (p=0.153). Comparison of the groups is displayed in Table 21 below demonstrates how the groups were dissimilar in all but their log mean warrants served/day.

Table 21: ANOVA Comparison among Intervention III Groups

		Sum of Squares	Df	Mean Square	F	Sig.
Log Arrests/Day 10-12_2000	Between Groups	.310	3	.103	6.782	**.000**
	Within Groups	4.565	300	.015		
	Total	4.875	303			
Log Arrests/Day 10-12_2001	Between Groups	.364	3	.121	6.904	**.000**
	Within Groups	5.239	298	.018		
	Total	5.603	301			
Log Warrants/Day 10-12_2000	Between Groups	.018	3	.006	1.769	.153
	Within Groups	1.018	297	.003		
	Total	1.037	300			
Log Warrants/ Day 10-12_2001	Between Groups	.060	3	.020	4.446	**.004**
	Within Groups	1.351	300	.005		
	Total	1.411	303			
Log Queries/day 10-12_2000	Between Groups	130.637	3	43.546	27.805	**.000**
	Within Groups	469.836	300	1.566		
	Total	600.472	303			
Log Queries/day 10-12_2001	Between Groups	75.412	3	25.137	15.860	**.000**
	Within Groups	472.323	298	1.585		
	Total	547.735	301			
Log Types of Queries 10-12_2000	Between Groups	20.968	3	6.989	39.786	**.000**
	Within Groups	52.702	300	.176		
	Total	73.670	303			
Log Types of Queries 10-12_2001	Between Groups	8.532	3	2.844	13.107	**.000**
	Within Groups	65.092	300	.217		
	Total	73.624	303			

Post Hoc tests are necessary to discern group differences (or similarities). Tests for homogeneous subsets (Tukey B and Scheffe analysis) confirm that log mean warrants per day pre Intervention III is one subset and not statistically different among any of the groups (p=0.426)(see Table 22).

Table 22: Post Hoc Analysis of Log Mean Warrants served per Day 10-12_2000

	Intervention III Group	N	Mean	
				Subset for alpha = .05
Tukey B(a,b)	Group H	12	.0291	
	Group E	124	.0399	
	Group F	136	.0540	
	Group G	29	.0544	
Scheffe(a,b)	Group H	12	.0291	
	Group E	124	.0399	
	Group F	136	.0540	
	Group G	29	.0544	
	Significance			**.426**

Means for groups in homogeneous subsets are displayed.

a Uses Harmonic Mean Sample Size = 30.022.

b The group sizes are unequal. The harmonic mean of the group sizes is used.

Further review of post hoc Bonferroni analyses demonstrate that in pre-intervention log mean arrests/day, Group F (replacing their computers with new ones p=**0.005**) and Group G (p=**0.001**) were statistically significantly higher performers than Group E. In post intervention log mean arrests/day, Group F statistically significantly outperformed Group E (p=**0.001**).

Table 23: Post Hoc (Bonferroni) Comparisons in Intervention III

Dependent Variable	(I) Experiment 3	(J) Experiment 3	Mean Difference (I-J)	Std. Error	**Sig.**	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
LogArrDay10-12/00	First time computer user in 2001	Replaced 1997 computer in 2001	.00836	.03729	1.000	-.0907	.1074
		1998 Computer upgraded w software	-.09329(*)	.02510	**.001**	-.1600	-.0266
		1998 computer replaced in 2001	-.05112(*)	.01526	**.005**	-.0917	-.0106
	Replaced 1997 computer in 2001	First time computer user in 2001	-.00836	.03729	1.000	-.1074	.0907
		1998 Computer upgraded w software	-.10165	.04214	.099	-.2136	.0103
		1998 computer replaced in 2001	-.05947	.03713	.661	-.1581	.0391
	1998 Computer upgraded w software	First time computer user in 2001	.09329(*)	.02510	**.001**	.0266	.1600
		Replaced 1997 computer in 2001	.10165	.04214	.099	-.0103	.2136
		1998 computer replaced in 2001	.04217	.02485	.544	-.0238	.1082
	1998 computer replaced in 2001	First time computer user in 2001	.05112(*)	.01526	**.005**	.0106	.0917
		Replaced 1997 computer in 2001	.05947	.03713	.661	-.0391	.1581
		1998 Computer upgraded w software	-.04217	.02485	.544	-.1082	.0238
LogArrDay10-12/01	First time computer user in 2001	Replaced 1997 computer in 2001	.03129	.04009	1.000	-.0752	.1378
		1998 Computer upgraded w software	-.07557(*)	.02698	.033	-.1472	-.0039
		1998 computer replaced in 2001	-.06346(*)	.01646	**.001**	-.1072	-.0197
	Replaced 1997 computer in 2001	First time computer user in 2001	-.03129	.04009	1.000	-.1378	.0752
		1998 Computer upgraded w software	-.10686	.04529	.114	-.2271	.0134
		1998 computer replaced in 2001	-.09475	.03993	.110	-.2008	.0113
	1998 Computer upgraded w software	First time computer user in 2001	.07557(*)	.02698	.033	.0039	.1472
		Replaced 1997 computer in 2001	.10686	.04529	.114	-.0134	.2271
		1998 computer replaced in 2001	.01211	.02675	1.000	-.0589	.0831
	1998 computer replaced in 2001	First time computer user in 2001	.06346(*)	.01646	**.001**	.0197	.1072
		Replaced 1997 computer in 2001	.09475	.03993	.110	-.0113	.2008
		1998 Computer upgraded w software	-.01211	.02675	1.000	-.0831	.0589

In log mean queries per day post intervention, post hoc Bonferroni group interval comparisons shows statistically significant differences. Group E (new users with mean=0.8518) is statistically significantly *lower* in performance than Group F (mean= 1.5832; p=**0.000**) or Group G. (mean=2.4345; p=**0.000**). Group G, in turn, is statistically significantly higher in log mean queries/day than Group F (p=**0.005**).

In types of queries post intervention, post hoc Bonferroni group interval comparisons also showed statistically significantly differences. Group E is statistically significantly lower in log mean types of queries than Group F (p=**0.000**) or Group G (p=**0.000**) as shown in Table 24.

Table 24: Post Hoc (Bonferroni) Log Types in Intervention III

Multiple Comparisons

Dependent Variable: LogTypes10_01
Bonferroni

(I) Experiment 3	(J) Experiment 3	Mean Difference (I-J)	Std. Error	Sig.	99% Confidence Interval Lower Bound	Upper Bound
First time computer user in 2001	Replaced 1997 computer in 2001	-.13384	.14082	1.000	-.5806	.3130
	1998 Computer upgraded w software	-.52719*	.09477	.000	-.8279	-.2265
	1998 computer replaced in 2001	-.25982*	.05764	.000	-.4427	-.0770
Replaced 1997 computer in 2001	First time computer user in 2001	.13384	.14082	1.000	-.3130	.5806
	1998 Computer upgraded w software	-.39335	.15910	.084	-.8981	.1114
	1998 computer replaced in 2001	-.12598	.14019	1.000	-.5708	.3188
1998 Computer upgraded w software	First time computer user in 2001	.52719*	.09477	.000	.2265	.8279
	Replaced 1997 computer in 2001	.39335	.15910	.084	-.1114	.8981
	1998 computer replaced in 2001	.26737	.09383	.028	-.0303	.5651
1998 computer replaced in 2001	First time computer user in 2001	.25982*	.05764	.000	.0770	.4427
	Replaced 1997 computer in 2001	.12598	.14019	1.000	-.3188	.5708
	1998 Computer upgraded w software	-.26737	.09383	.028	-.5651	.0303

*. The mean difference is significant at the .01 level.

Summary of among Group Testing by ANOVA and Post Hoc Tests

Intervention I -- The two intervention groups are statistically significantly different in their log mean arrests per day. Pre-intervention, Group B consists of much higher performers than the Group A mean performance. Post-intervention, Group A statistically significantly improved but still never reached the Group B performance level.

Intervention II -- The three intervention groups are statistically significantly different in their log mean arrests per day. The two groups using computers, Group B and C, are much higher performers, in log mean arrests/day than Group D, without computers and Group C are higher performers than Group B, using Intervention I computers, even though they were equivalent pre-intervention (note – during the period tested in *pre-Intervention II*, the Intervention I Group B, with computers, was equivalent in log mean arrests/day to the Intervention II Group C, *who were without computers at that time*).

Intervention III – All groups are statistically similar in log mean warrants served per day pre and post intervention but the performance similarity stops there. Group E (new computer users) are statistically significantly lower in performance through log mean arrests/day, pre-intervention, than either Group F (replacing Intervention II computers) or Group G (upgrading Intervention II computers). Post intervention, Group G is statistically significantly different, with higher performers, than Group E. Post intervention, Group E are statistically significantly lower in performance than either Group F (replacing Intervention II computers) or Group G (upgrading Intervention II computers) in log mean queries/day and types of queries. Group G is also statistically significantly higher in types of queries than Group F.

Performance output measures can easily be misconstrued however. It is quite possible to examine the mean number of queries for all wireless users (Table 25) and see the mean number of queries per officer per day, for the mean number of wireless users, has increased over the sampled mean of 6 queries per officer per day. The same can be said for the overall number of types of queries used above the "basic four." Research into group characteristics explains increases in outcome performance measures than examining cumulative group changes as follows.

<u>Total users on radio queries versus wireless show activity increases</u>

Changes across interventions appeared to be generating more and more queries as wireless users were added. Comparing total number of wireless queries to the numbers of all radioed queries gave the following chart to illustrate the quantity differences. Figure 28 below represents available data of total users of radio and wireless terminals beginning post-Intervention II. Note that users continued to be added after September 2001 but that was beyond the study period.

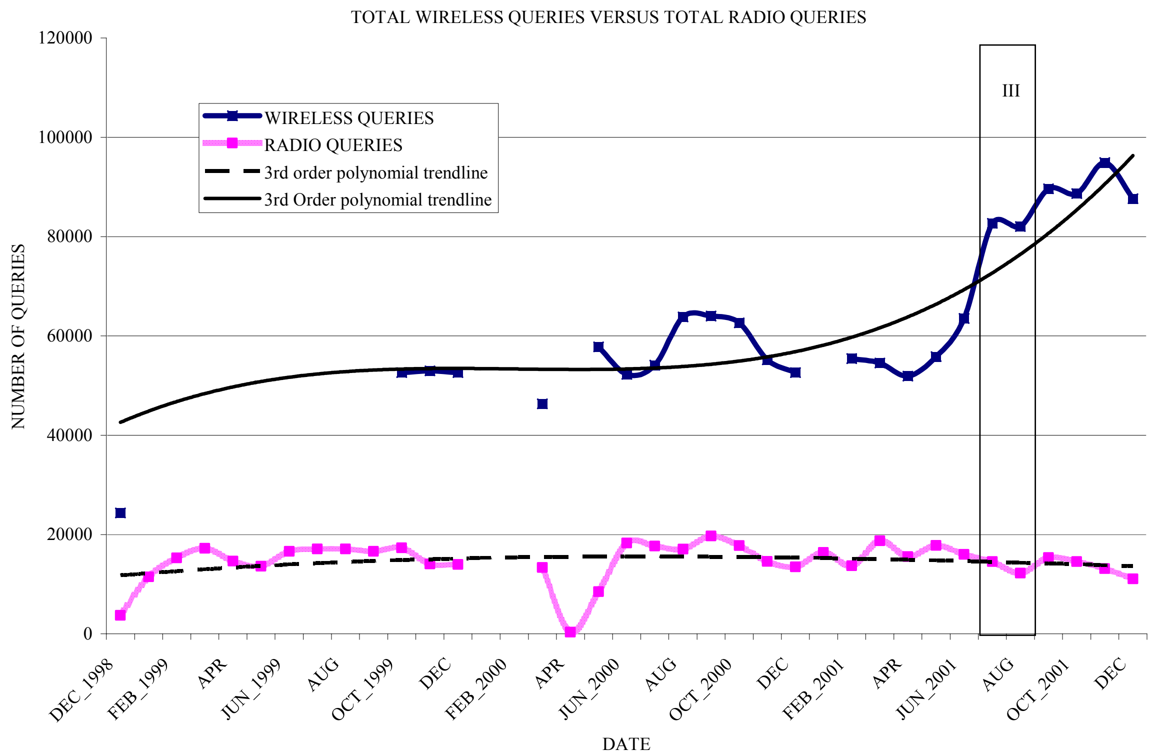

Figure 28: Post Intervention II Total Wireless Vs Radio Queries

Intervention III is indicated by the shaded area on the graph above as the figure only shows the period post-Intervention II. These data cannot be used to show direct effects for any intervention but there were statistically significant differences between the means for all wireless queries per month versus total radioed queries/month.

It was evident that wireless users were, on average, using more types of queries than the "basic four" used in radio queries and a greater number of queries than in mean queries by radio. Officers performing radio queries could have used any queries similar to wireless users. Queries *by radio* stayed relatively consistent when observing available data from December 1998 to December 2001 (mobile computer queries are not available for January and February 2000 or April 2000 and radio queries are not available for January and February 2000).

Table 25 below describes aggregate data of all wireless users and all radio users over thirty-seven (37) months of available records for the study period. "N" represents the months of available data for all wireless and all radio users.

Table 25: Descriptive Statistics for # Queries, & Types of Queries for All Users of Wireless and All Radio Query Users per month

	Min	Max	Mean	Std. Deviation	N Months
TOTAL_WIRELESS USERS	166	514	**274.81**	122.749	26
TYPES_WIRELESS_QUERIES	56	80	**66.23**	5.948	26
TOTAL_WIRELESS_QUERIES	24372	94880	**62420.58**	16828.859	24
TOTAL_RADIO QUERIES	0	19687	**13747.24**	5075.118	37
Valid N (listwise)					24

The total types of queries used per month Figure 29 (below) displays the number of query types accessed by all wireless users from December 1998 (post Intervention II) to December 2001 (study end) in monthly segments. Wireless users averaged 66 types of queries used versus four types (average) used by officers using radio (refer to Table 4).

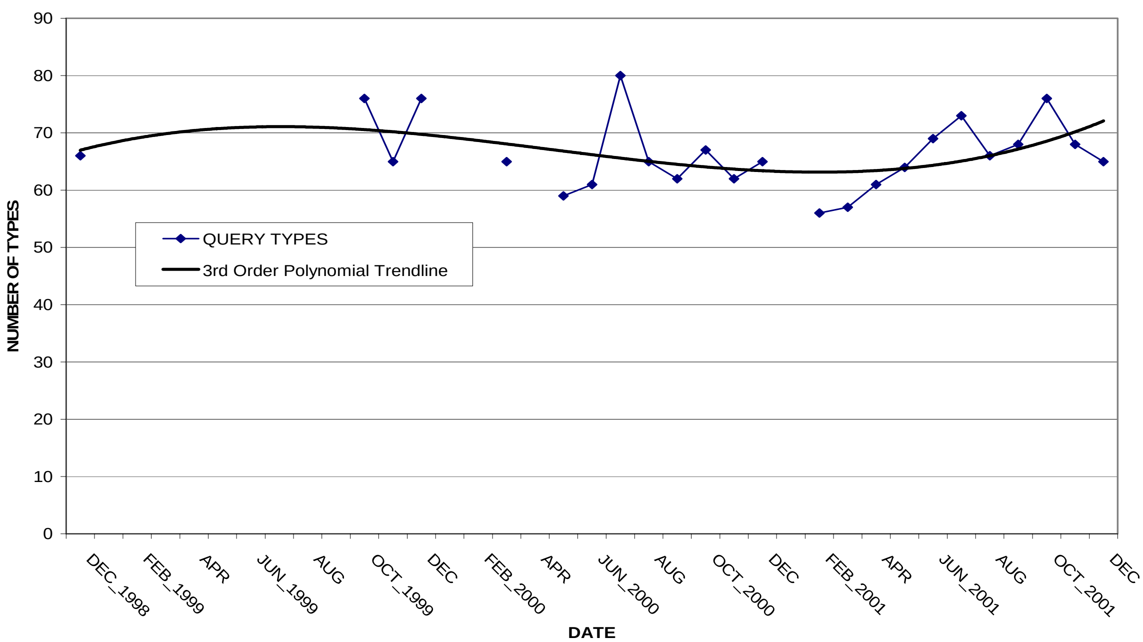

Figure 29: Types of Queries by All Wireless Users

A comparison of total wireless user queries mean versus total radio queries mean with a large sample significance test (Agresti and Finlay 1997).

$$\hat{\sigma}\bar{Y}_2 - \bar{Y}_1 = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\hat{\sigma}\bar{Y}_2 - \bar{Y}_1 = \sqrt{\frac{16829^2}{24} + \frac{5075^2}{37}}$$

$$\hat{\sigma}\bar{Y}_2 - \bar{Y}_1 = 3535$$

$$z = \frac{\bar{Y}_2 - \bar{Y}_1}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$z = \frac{62420 - 13747}{3535} = 13.77$$

$$p \cong 0.000$$

Where:

- n_1= 24 months for the number of months of wireless data
- n_2= 37 months is the number of months of radio query data (from Table 24)
- s_1= 16,829 for the standard deviation in wireless queries
- s_2= 5,075 for the standard deviation in the number of radio queries during the months of collected data from Table 24

This statistically significant difference demonstrates that additional wireless users have more queries than radio users. The radio queries may not be reduced by adding wireless queries because the radio query system was so capacitated that even adding large numbers of wireless users did not relieve the congestion even though all patrol users received wireless computers by December 2001 (see Figure 28).

There were slight increases in mean numbers of queries per day in every group in Intervention III though they were not statistically significant across the intervention for each of the groups (E, F, G, and H); their cumulative increase was almost 9,000 additional queries per month (from mean numbers of queries per user in Figure 28). The standard deviation for the radio queries was 5,075 queries per month (from Table 24). The groups were busier but not necessarily more productive according to police performance standards.

The key finding from this aggregate analysis of all user performance is that the cumulative records can be a poor measure of performance and generally does not define where performance changes occur or who is doing them. Poorly performing groups could easily mask true performance gains by other groups leading to a productivity paradox.

CHAPTER VI

DISCUSSION AND CONCLUSIONS

Groups explain the results

The overall increase in total queries would lead management to conclude that the entire base of users was improving when that was not the case. These figures could easily mislead management. Managers looking at the last two aggregate graphs would believe they might be able to improve the performance of low performers in an organization by giving them fast, and expensive, new tools. This was not the case in this study as the low performers never improved, even marginally when using the new equipment as there were no changes in arrests or warrants served with Group E (30.5% of the total officers in the study).

The hypotheses summary appears contradictory until the implications of all interventions are considered as well as the overall group effects on these hypotheses. Hypotheses were supported for Groups A and F but not for other groups. The apparent contra-intuitive results of their performance outcomes are as follows:

There are statistically significant increases in the number of queries with reductions in latency (Hypothesis$_{1A}$) with Group B in Intervention I but there is no corresponding statistically significant change in their performance measures (mean arrests/day and mean warrants served/day) across Intervention I. There is a statistically significant change, however, in performance for the non-targeted Group A, those *without computers*. The non-computer group A was not shown to have a change in information latency yet they, the non-targeted group, showed significant gains in both mean arrests/day and mean warrants served/day while the group with computers showed no significant changes. I suspected that some in Group A were radioing Group B officers to run queries for Group A instead of radioing records' clerks.

With Intervention II, Group B, using Intervention I computers, experienced no change in information latency and no statistically significant changes in mean arrests/day or mean warrants served per day. Group D, without computers showed no statistically significant change in mean arrests per day or mean warrants served/day as had Group A during the earlier intervention. The 192 officers in Group C, using Intervention II computers experienced a significant, across group comparison increase in mean arrests per day but not in mean warrants served/day. The same possible change in latency in Intervention I had caused Group A, without computers, to improve in both mean arrests/day and mean warrants served/day. There was no similar increase with Group D even though Group D is a subset of Group A. Group D, without computers in Intervention II, would have been expected to have better performance measures after the intervention if they had acted like the high performers from Group A. They did not improve.

The officers' performance in Group C improved when they got their own computers even though Group C was a subset of Group A also. Obviously Group D and C, both subsets of A, were different in performance traits but the officers chosen to be in Group C were higher performers than the rest of Group A. Further, queries per day changes show evidence of these differences. I have already shown that group C, in Intervention II, were the high performers from Group A in Intervention I. Group C had a mean of 12.68 queries per day after Intervention II. Group B demonstrated 19.2 queries per day immediately after Intervention I but averaged a mean of 11.14 queries per day, a year after Intervention II. Their mean dropped but this was not a statistically significant drop, in paired t-tests, because their standard deviation was large and the group small (n=34). It is clear that computer assignments to personnel are not random in this study because of the group differences.

Group E officers in Intervention III experienced a large reduction in latency which did not cause any statistically significant change by mean queries per day or mean types of queries. There were also no statistically significant changes in mean arrests/day or mean warrants served/day for Group E. Group F, however, experienced a moderate but statistically significant increase in latency over the intervention when going to the new computers (the computers were faster but its network data throughput speed was slower than the system it replaced) and showed a significant increase in the types of queries. Queries for this group did not statistically significantly increase but they displayed a statistically significant decrease in mean arrests/day. They performed a proportionate number of queries per day, post intervention versus pre-intervention, even though there was more latency. Officers might have been unable to complete arrests in the time allotted since the same number of queries would have taken this group twice as long to perform. This would mean that it was not information latency, in this case, affecting the number of queries as much as **the increase in information latency affecting the decision process** for arrests. This would require a model revision in which information latency moderates the decision-making process. This could be diagrammed as information latency moderating the Recognition-Primed Decision Process as shown in figure 30 below.

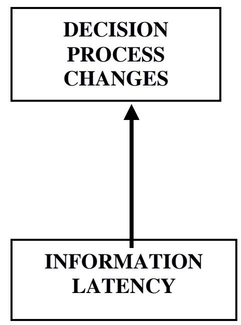

Figure 30: Information latency moderates decision process changes

<u>Tracking the apparent anomalies</u>

While the data is not available, the effect, which occurred across intervention I, is clear if it is caused by a high number of radio queries by the group without computers, Group A, to the officers with computers (Group B). The comparison of groups in Intervention I demonstrated that the groups were dissimilar in performance measures related to arrests. The groups were similar in age and experience as was shown earlier however. The Intervention I groups were dissimilar in performance as it impacts arrest performance based on elements of crime but not statistically dissimilar in arrests based on identification of subjects. A reevaluation of the high performers in Group A showed they lost no time exploiting ways to avoid information latency by radioing the officers with wireless computers instead of waiting in line to radio the remote records clerks. This would suggest that high performers try to avoid latency to achieve better performance outcomes.

High Performer Latency Avoidance

<u>User performance traits affecting outcomes</u>

There were indications of officers circumventing the records latency bottleneck by radioing other officers in their geographical area who had wireless laptops asking them to run queries for them. There were anecdotal reports that Group A officers had radioed Group B officers frequently after Intervention I but there was no documentation on the number of queries from Group A. These radio queries from other officers in the field would have greatly increased the workload for these few officers with computers per shift, Group B, and removed a large queuing latency of any radio query. Group A officers radioed colleagues with wireless computers instead of radioing queries to clerks. They could stay on the same radio channel and face no wait because of the officers ahead of them when performing a radio query through records. This is similar to going from one checkout clerk at a grocery to several clerks when the store gets busy (refer to Figure 31 below).

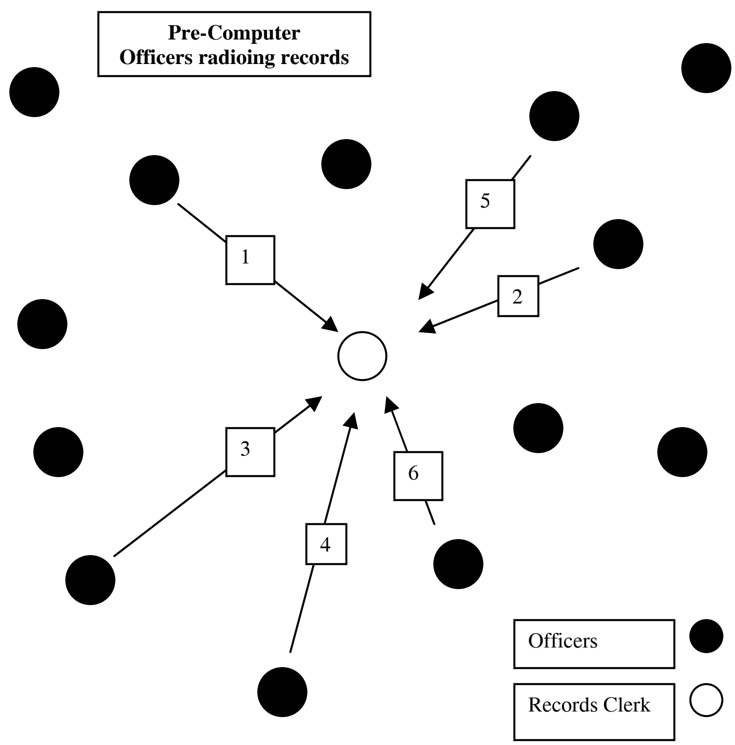

Figure 31: Pre-Computer Radio Queries

In 1997 there were approximately 78 officers working per eight hour shift (Planning and Research data) out of the 500 assigned to patrol. With 34 laptops distributed throughout the county on three different shifts, there were normally 1 to 2 wireless computers in use per sector (geographical boundary of coverage now called precincts). There was still only one records terminal for the entire county. The latency avoidance after Intervention I is clarified in Figure 32 (below).

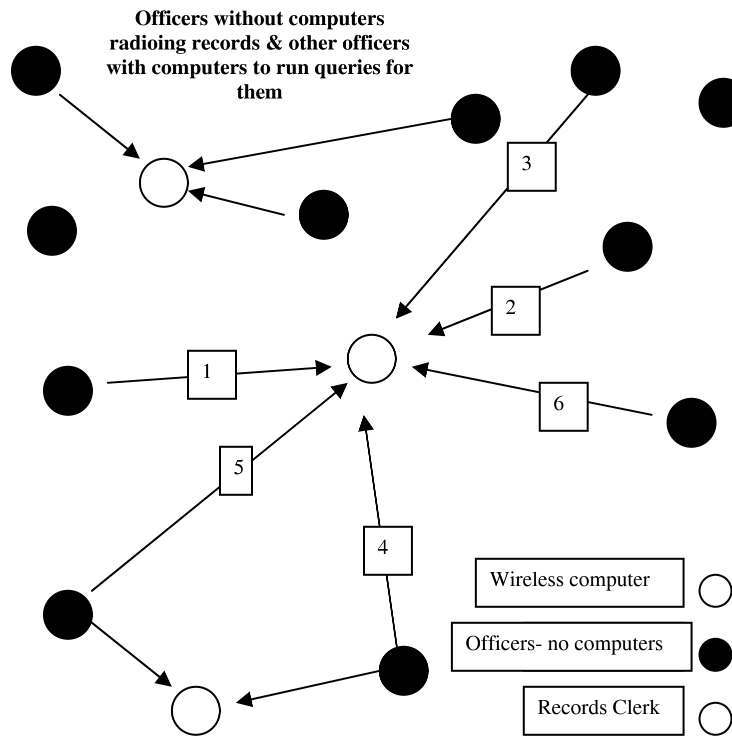

Figure 32: Post Intervention I Queries

There were still many officers in patrol who would radio queries through records' clerks. The high performing officers in Group A, apparently frequently called Group B officers due to the Group B's large numbers of queries. This would increase workload for the Group B officers with computers, by increased radio queries from other officers. This certainly explains how officers in Group A, without computers, improved in performance as they would have experienced a large decrease in latency and would have received information from Group B officers who have with significantly more training than the average civilian clerk. Group analysis already demonstrated that the high performers in Intervention II Group C were the high performer subset of Group A.

If this was indeed the case then adding 129 more wireless computers during Intervention II would be affected greatly by who got new computers, by their performance traits, as much as any latency changes from the computers themselves. In fact, ANOVA and Post Hoc analysis among groups *pre* Intervention II showed that the officers getting new computers in Intervention II had performance characteristics similar to Group B who were already using computers. These officers in Group C were the high performers from Group A, without computers in Intervention I. Group C equaled Group B in pre-Intervention II mean arrests/day because Group C were the 'high performers' from Group A performing radio queries through Group B officers.

While there were 129 computers issued, there are 192 computer users listed for Intervention II computers because different officers used the same Intervention II computers (identified through a unique terminal ID number in departmental data) on different shifts. Data showed that officers in the "no computer" group, post Intervention II, borrowed computers, after a supervisor's approval, from one shift to the next. State and federal rules mandate strict login criteria and all queries from any wireless terminal are identified by terminal identification ("pid number") and employee.

High performers would shift the radio query process in patrol as suggested in Figure 33. This network diagram of post Intervention II helps demonstrate the query flow from users. Grey circles are the officers with wireless computers performing queries. Some of the officers without computers would radio queries to records, other officers or borrow a computer themselves.

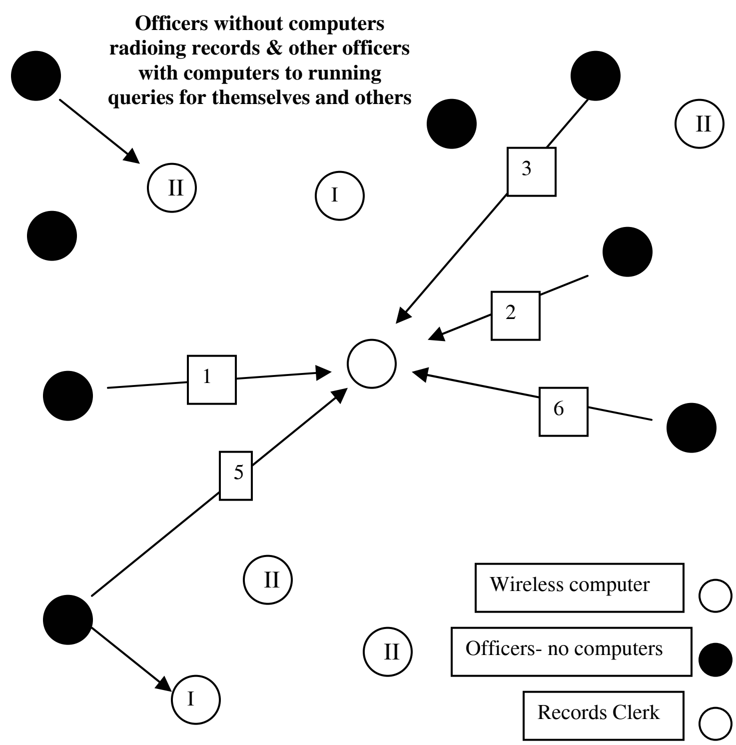

Figure 33: Post Intervention II Queries

There are approximately 20 to 25 wireless laptops deployed at any time post Intervention II versus only 7-8 wireless laptops deployed and in use at any time during Intervention I. This would mean that the wireless computers deployed in police use would have been 4 or 5 per shift in each geographical sector (approximately 16 officers would be on duty per shift at each geographical sector or precinct). More laptops in the field meant that more users could run their own queries. The high performers in Group C could also now reduce their latency almost two minutes per query since they would not have the call someone else to run their queries for them.

Another high performer latency avoidance effect

Another variable, a possible history effect, is that officers after Intervention II (early 1999), discovered how to write macros and set up buttons on their computers that would type the esoteric query prefixes (examples: QEXID SSN/ for social security checks or QTOM OLN/ for queries of moving traffic offenses from an in-state driver's license) before the officer typed the query body. This is to overcome a portion of the difficulty of typing in low light conditions (dusk to dawn) as the computers did not have lighted keyboards until Intervention III computers were issued. Officers reduced latency in typing the initial query prefixes with this technique but unfortunately, such timing data was not captured. The group C officers shared these button templates by magnetic media throughout the department.

The high performers drawn from the "no computer" Group A in Intervention I -- obtained use of computers during Intervention II. They made use of them by significantly increasing mean arrests per day when their latency decreased significantly. In Intervention I, the high performers in the "no computer" group reduced their latency and increased their ability to perform more queries by eliminating queuing times of 5 minutes per query by simply radioing the officers who had already been issued wireless computers. This practice was reported during the period 1997 to early 1999 from information shared with Planning and Research laptop support staff and training academy instructors but there is no recorded evidence as the radioed queries were often car-to-car (unrecorded) modes to officers in neighboring zones.

User performance traits affecting Intervention III

The 123 officers getting computers issued for the first time (Group E) were, pre Intervention III, lower performers in arrests/day than either Group F (replacing Intervention II computers) or Group G (upgrading Intervention II computers). Group E, unaffected by the technology, were lower in arrest performance than either Group F (replacing Intervention II computers) or Group G (upgrading Intervention II computers) in both numbers of queries and numbers of types of queries. These 'unaffected' officers were the officers remaining after Group C got computers in Intervention II (where Group E is a subset of Group D which is a subset of Group A). The unaffected had lost their high performers in the first two interventions. Now, with an opportunity to perform wireless queries, they performed as if the computer was never in place and showed no significant improvement in any area.

Group E experienced a significant drop in information latency when going from radioed queries to wireless queries yet post-Intervention III had roughly the same number of queries as anticipated with radio queries. Group E also showed an average number of types of queries that was equivalent to the radio queries typically used by personnel without computers. Group G used significantly higher numbers and types of queries than Group F in pre and Post Intervention III. Group G, the better of the best, kept the computers with the lowest latency, post Intervention III.

Groups F and G experienced slightly higher latency but only Group F showed a statistically significant increase in types of queries. This may be because 63 of the 138 officers (46%) getting new computers in Group F had been using borrowed computers after Intervention II and this was their first chance customize their personally issued computer. They would not have customized the borrowed computer as query buttons were normally customized only by the persons actually issued the computers. Group E simply used their wireless computers in the same way they had their radios, with similar performance.

Any model would have to differentiate between the low performers, unaffected by the technology, and the high performers who may change decision-processes to improve performance. User performance characteristics should be identified prior to technology deployments in other arenas of high-risk decision-making (Figure 34).

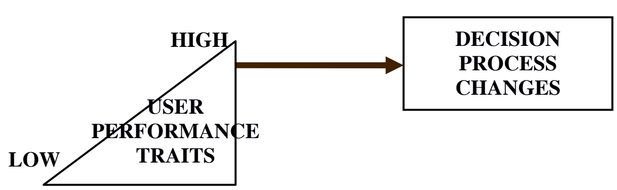

Figure 34: User performance traits are mediated by the decision-making process

Spelling out high performer latency avoidance

It seems clear that low performers in Group E are not improved with large decreases in information latency. The large drops in information latency (i.e., increases in speed) did nothing to increase the number or types of queries over what was available through radio queries or increase their mean arrests/day or mean warrants served/day.

Conversely, high performing individuals appeared to go to great lengths to reduce latency and continue to maintain high performance outcome measures. They will circumvent traditional processes to improve their, already high, performance. Significant reduction in information latency adds to their likelihood of producing higher outcome measures (*i.e.* Group C mean arrests/day). Moderate increases in latency may impact their performance. Particularly, a significant increase in types of queries resulted in a significant decrease in mean arrests per day. This may mean that using multiple cues was simply the result of 63 of the 137 officers in Group F getting their own computers for the first time and adapting them for additional queries with customized macro buttons. They would not be allowed to customize any borrowed computer. Slight changes in latency do not significantly change performance.

Information latency and crisis information systems

Information latency appears to affect high performers but it took significant reductions in latency to improve performance. Reductions characterized in minutes meant more than reductions characterized in seconds but only with high performers. Slight changes in latency have no effect (*i.e.* Group G). Group F experienced a moderate increase in latency that was statistically significant. They performed approximately the same number of queries as when they had their faster system but this same number of queries would have taken twice as long to perform. Their arrest performance dropped dramatically. This increase in time for queries could easily impact how much information they were accustomed to getting in a given time. Now, with slower access, officers may have had to release persons due to legal time constraints.

It may be that systems that rely on time pressured, critical information, as with crisis information systems, may be affected more by the increases in information latency than by decreases in information latency. It may also mean that significant drops in latency with no corresponding improvement in performance may be related to low performing users instead of the system itself. High performers may be more affected by any latency changes as is illustrated in an example from this study.

CHAPTER VI

CONTRIBUTION

Differentiating high and low performers

User performance traits should be differentiated when introducing a technology. High performers, for the purposes of this study, are those people who have high performance statistics compared to others in their workplace. High performers might be identified from the upper quartiles using whatever productivity measures an organization already uses. Lower performers in an organization would be defined by the lower quartiles in whatever performance measures already used by the organization. Low performers, in this study Group E, used the new technology as a simple substitute for their past system and thus do not accrue benefits of any technical advantages. Low performers do not appear affected by getting information faster in high-risk decision making. They did not seize the advantage as did the high performers. The introduction of new technology to high performers first should have performance advantages over a random introduction of technology to a group with mixed performance characteristics.

User performance traits more important than information latency

Performance traits of the decision-makers appear to be more of a critical factor in high-risk decision-making than information latency. Results suggest that significant increases in the number of queries may affect decision thresholds *in high performers*. This result would infer that additional queries (*i.e.* more information) aid in clarifying anomalies when a high-performer diagnoses a high-risk situation. This lends support for *more data* in the diagnostic portion of the recognition-primed decision (RPD) model but *only for high performers*.

An increase in multiple cues, defined as different types of queries in this study, are supported in this study as contributing to *negative* performance with Group F. It is not clear how this might significantly reduce arrests/day but it appears that the same number of queries, pre and post intervention, for this group took twice as long. The increase in latency had more effect on this drop in performance than an increase in multiple cues. Again, performance changes that support how immediate feedback and multiple cues reduce uncertainty in the media richness model worked *only with high performers*.

There also appears to be a threshold effect with information but *only where it pertains to high performers*. Low performers did not increase information gathering in high-risk decision-making even when presented with large scale decrease in information latency. Large scale decreases in latency with high performers permitted increases in performance.

<u>Theoretical contributions to the RPD Model from this study</u>

More queries did lead to better performance in this study, giving credence to the RPD model portion that *more data* clarified more anomalies in high-risk decision-making. The *actions taken,* directly affected performance outcomes, but *only with high performers*. The RPD model may need to be adapted to differentiate high performers or on user performance traits. It seems clear that the higher performers were using decreased latency to increase the number and types of queries to clear up anomalies in arrestees' stories. It may simply be necessary to explain that the RDP model is for experienced individuals who already perform above the norm.

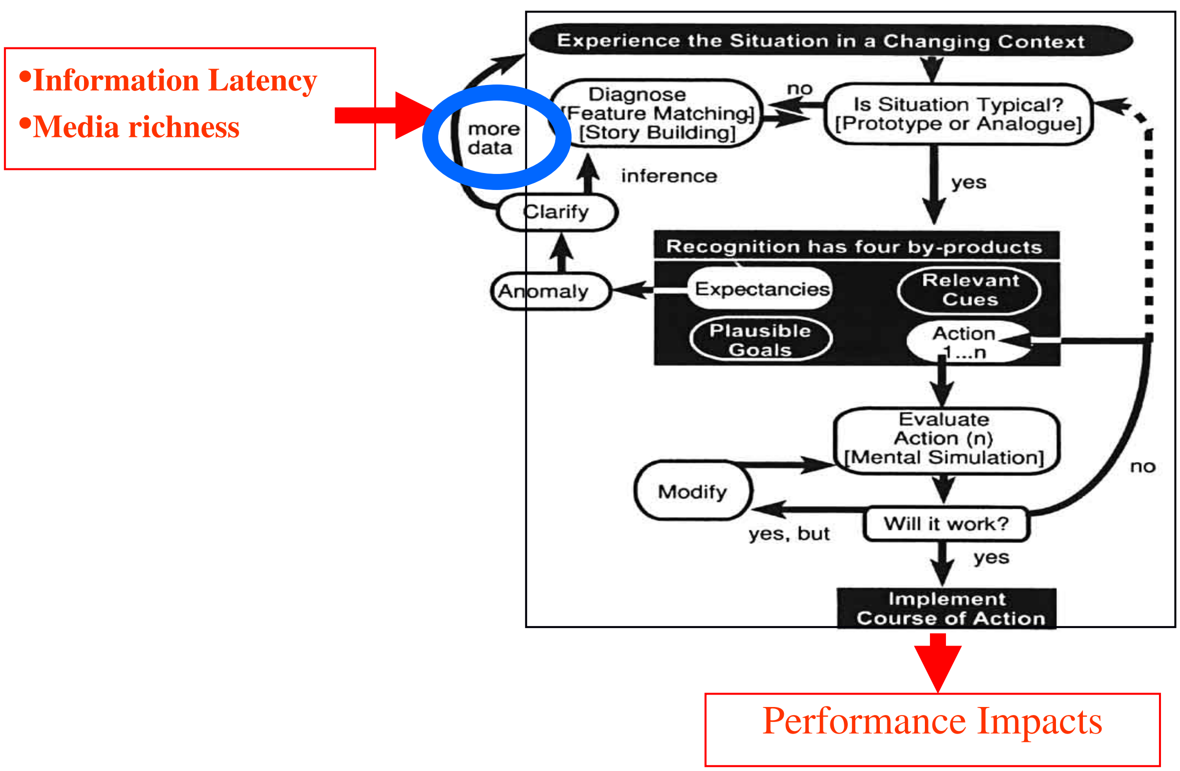

Figure 35: Extending the RPD Model with user traits and performance

Theoretical implications to media richness validation

Large scale changes in feedback, with high performers, affected performance also. The fact that user performance characteristics directly affected whether significant changes in feedback or multiple cues had any impact on performance may well explain why media richness has had mixed validation (Rice 1992; Dennis and Kinney 1998; Dennis and Valacich 1999; Kahai and Cooper 2003). The high performers certainly used feedback and multiple cues to reduce equivocality. How can one say an effect was not achieved by changes in media richness if the performance traits of the participants in any study are not considered? Data from past media richness studies might be reinvestigated, after the separation of high and low performers, to make sure low performers did not dominate any samples. Even high performers did not change performance when there were not significant changes in either feedback or multiple cues.

Theoretical implications of threshold effects

Thresholds may affect performance, with high performers as with Group F, moderate to significant changes may have to be present before any effect is measurable. Moderate changes in latency, feedback, and multiple cues did not significantly impact high-risk decision making with any of the three interventions with low performers even when they received moderate but significant drops in latency during Intervention III and performed no differently than with their pre-intervention utilization of radio queries. Media richness or recognition-primed decision models, in the same way, might not demonstrate any effect on high-risk decision-making performance unless high performers are evaluated and technological changes were significant. A threshold between old and new techniques or technology had to be moderately different in this study before discernable changes occurred.

Thresholds appear to moderate information latency if there are significant changes, with high performers. Group F had significant drops in mean arrests per day after significant increases in latency. In Group A, it appears that the high performers in that group were responsible for the significant increase in mean arrests per day by having officers with computers query for them.

The productivity paradox and predicting general user performance from a test group

It would be a serious mistake then to predict an overall performance change after using only one test group because of the probability of skewed group performance trait differences. It is reasonable to select test group individuals from high performers, low performers and average performers and test any new technology treatment with all groups. This is an alternative to predicting sweeping changes from the first test group of mixed performers where there might be a preponderance of high or low performers. High performer latency avoidance may also account for many programs being deemed an overall success when it was the high performers who made the total group's performance improve. This may explain the computer productivity paradox where large increases in spending may or may not be demonstrate measurable improvements in user performance (Brynjolfsson and Hitt 2001). Similar predictions could have been made in this research for overall productivity gains if these interventions had not been further investigated.

All groups are not equal

Performance improvements for an entire group cannot be generalized from improvements in one subgroup. In each of the three historical interventions in this study, the groups were indistinguishable, statistically, from each other in age and experience. However, work habits and performance traits, of the groups differed. The same hardware was given to several groups but they did not all fare identically with their performance metrics. Certainly, information latency affected the high performers or they would not have gone to such lengths to avoid bottlenecks in Intervention I and II. However, it took moderate to high changes in information latency before any statistically significant effect was measured.

This has broad-reaching impact on computer system evaluations or predictions of performance. No longer can one say that because such a system is so much faster, more work can be done. It is the user performance traits that upstage changes in a system. In this study, small increases in speed through information latency reduction had no impact on user performance and moderate changes were necessary before significant user performance changes were noted. This has practical implications for its implementation.

Results impact system purchases

This study has been written in as "technology neutral" way as possible to concentrate on the impact of information latency. The latest system used in Intervention III had the slowest wireless query response times (moderate latency) of any of the three interventions because of the capacitated system (due to congested, trunked radio transmissions). This ability to display information was still ten times faster than radio queries but twice as long as the "older" computers. Any significant changes in information latency due to system communication access should be a major consideration when considering a "system" because high performers will be affected.

The fast computers in Intervention III had difficulty getting their information through "cocktail straw" data access network. It would have been similar to going from fast cable modem access to dial-up Internet access. Managers should take into account the acceptable levels of information latency in critical systems in any "system planning" and not use the heuristic 'faster computers are better' with any replacements or upgrades. Managers should regularly examine system capacity to determine if distributed access might relieve 'bottlenecks' as were evident with the radio queries to remote clerks in this study.

Further impacts on resource allocation

This research also has impact when there are limited resources available to a group, company or entity. Managers wishing to optimize their results should target their limited to resources to the highest performers in their organization. They should only add, replace or upgrade when that replacement is likely to give significant increases in displayed performance. This might mean that existing hardware infrastructure could be used while investing in networking access upgrades. Higher speed data access might provide wholesale upgrades in performance using the same computers or the same computers with faster information display from upgrades in video cards. This also could be applied to decision support systems or command and control systems.

Changing low performers to high performers by compiling differences

If large scale system improvements had little effect with low performers; how do you change low performers into high performers? It may be that techniques culled from the high performers may be gathered and made accessible to the middle or low performers as was done in a study with Xerox service technicians who pooled techniques that proved effective (Raiman 1990). The Xerox database, made initially without management approval, used tips and techniques garnered from technicians who had solved installation and repair problems in the field. Techniques that proved effective were retained while techniques that were not effective were culled. After two months, the technicians using the database showed a ten per cent improvement in lower costs and lower service time than the technicians without the system. Similar differences might have been used in this case with the wireless laptops.

<u>Story of an outlier</u>

Officer D was a forty-year old making 300 arrests a year in 1998. He got the wireless laptop and the three hours of training that went with it. His arrests went up to 900 arrests per year which amounts to almost four arrests per day. He never got complaints and he made cases that were never thrown out in court. How did he do it and what did he do different from his fellows? He was able to use all the features of the laptop and natural interview skills. He was not a computer expert. Upon questioning Officer D, he stated he just asked questions in a different order than a suspect would expect and he "tied the information together." Instead of asking name, date of birth and social security number he might start out with a question, "Where did you go to high school?" and "When did you graduate?" He then would approximate their age and they would often lie and he would then concentrate on the anomalies in their story. If he found an active warrant on file, which was often, he would place them under arrest. The person arrested was often driving a car and he would ask if there was a friend of the arrestee he could telephone to drive the car home for the arrestee instead of towing it. Officer D would call the "friend" and ask if they would drive the arrestee's car home and did they have a valid driver's license. The "friend" would give Officer D their operator license number and while they were coming to pick up the car; Officer D would check the "friend" for warrants. He would often end up with a car load of arrestees.

<u>Training others users with the outlier's methods</u>

The project manager found the outlier's methods so compelling and useful that a training video was proposed to train all users during Intervention III. A local television station was approached to produce the video and the station offered to edit a film the video. A seventeen minute video was produced using Officer D with supplemental handouts and given to the police Information Services department for use while training wireless computer users.

Minimal technical proficiency stressed instead of operational use

The videotape and supplemental material was never used because only minimal technical proficiency in operating the wireless computers was stressed by the information services instructors. No actual field officers, experienced users, were allowed to train other officers. Wireless computers were supported by a civilian support staff and training had to be approved by the information services department. The years of experience in field use of this technology was not allowed to be disseminated to new users. The civilian staff felt uncomfortable teaching new field interview techniques designed for the technology. Later, in-service classes were not designed for hands-on training that was necessary when new computers were first issued.

Failure to address difference in user habits may have contributed to less productivity

Users differed in performance. The fact that high performer techniques were not compiled and used to train lower performers in Intervention III appears to be a wasted opportunity by management. Videotapes made by the highest performers in the police study with the wireless laptops were never used in training any classes because only minimal technical competency was stressed. Management concentrated on the minimal training to operate the technology and did not concentrate on optimizing the advantages of the new technology. If new wireless users use their technology simply as a substitute for their radio, they did not accrue the advantages of that technology. It seems clear that different media offer advantages or disadvantages and not everyone is able or willing to discern these differences or apply them to their daily routine. Differentiation of high and low performers in an organization is a logical first step to optimal use of a technology. It might have been possible to improve the performance of the low performers with training if performance characteristics had been ascertained prior to deployment.

Assumptions and Limitations of the Study

False positives in decisions are not captured in this study since data where a magistrate refuses giving a warrant to an officer or misidentification of a subject occurs are not captured on reports due to the exceedingly low probability of occurrence. Numbers of officers did not regularly change geographical sectors or outside precinct boundaries as this was not the general practice, though officers were removed from the study when they transferred to departments other than patrol. Bike and walking officers did not regularly use a computer unless they shared one in a car during inclement weather.

Subjects were removed from the study due to transfer, retirement, resignation, termination of employment or if they had disciplinary suspensions. Large sample sizes and pair-wise elimination across interventions were used to give a reliable sample. The individual circumstances of each arrest were not characterized in this study or even the number of stops. Data was not kept on *stops with no arrest*, where an officer stops a person and then determines to release that subject because criteria for arrest are not met. It was only in 2003 that such data began to be collected. This means that *more data* cannot show any increase in the number of stops, but only stops where arrests are made.

The data collected on queries also does not show whether particular queries to check for active warrants received responses that are positive or negative for warrants on file. It can be determined however, that particular officers send a specified number of queries, of different types, and then make arrests of a particular nature. Outcome measures were recorded since the initiation of this archival data study, pre-intervention but the number of pre-intervention queries cannot be obtained prior to December 1998. Changes in performance measures must be inferred by individual changes using numbers of pre and post intervention observations. It is also an issue that data collected needed to be disaggregated into monthly segments instead of quarterly blocks to show individual officer differences. This research did not examine the effects of study participants gaining rank or gender effects and those aspects should be examined in future research. Further, only a single organization was studied.

A new user-performance model

User performance traits directly affect performance outcome measures in high-risk decision making. Information latency does not drive performance but may moderate performance through its effect on the decision-making process in high-risk decisions. Information latency moderated the performance of individuals with high performance indexes, compared to others in their organization, making high-risk decisions by driving the level of feedback and multiple cues. A threshold had to be reached before significant effects in the high-performers could be reached. Adding user performance traits to the "*Is situation typical?*" section and performance outcomes after "*implement a course of action*" in the RPD model may better model high risk decisions. The model below is suggested by the results of the study (see Figure 36):

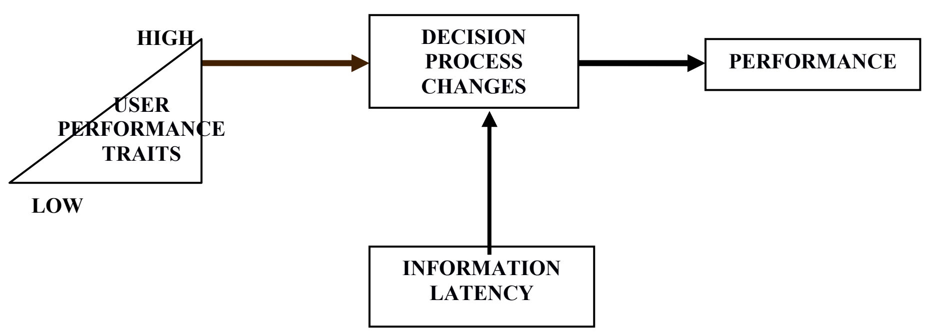

Figure 36: Proposed User Performance Model

Information latency is important but only in a secondary vein. The user performance traits were the key factor in how the users adapted to the new technology. There appear to be process changes where users avoided any built-in or perceived information latency. Any information latency avoidance is important. Examples include when the users would borrow computers (data from 63 out of 192 in Intervention II) or where officers wrote their own query macros after Intervention II to customize the types of queries they would use. They also may have avoided information latency by calling officers with computers on the radio to avoid queuing time as evidenced by reductions in the mean number of queries that occurred for Group B when group C obtained their own computers. Information latency appears to moderate the decision process. These changes in decision processes are the mediating variable between user performance traits and performance changes.

In other words, it is the user performance traits that 'drive' the decision processes with information latency affecting how long it takes users to get that information. Significant changes in information latency are necessary before performance changes are evident. Performance traits drive the decision process with latency changes aiding or hindering that endeavor.

Future research

The fact that user performance played such a part in this research does not mean that the concept is immediately transferable to other domains but this surely should be investigated. Any research measuring user performance changes should first examine group performance traits instead of simply looking for changes across a treatment. The research should further understanding of real-world decision-making during high-risk situations that may be generalized to other high-risk situations (emergency management, military, medical emergencies, law enforcement, firefighting) where, *high-performing*, experienced decision-makers are working with limited facts in chaotic situations. The RPD model should be reassessed, after adding user-performance traits (Klein 1989; Roberts and Dotterway 1995; Kaempf, Klein et al. 1996; Miao, Zacharias et al. 1997).

The media richness model should be further scrutinized after determining the initial performance levels of study participants since initial user performance traits may determine outcomes more than media richness itself. Decision thresholds should also be considered when considering system requirements and managers should consider establishing capture of performance cues for easy program evaluation at any stage (Gilliland, Benson et al. 1998; Weaver and Richardson 2002). Small changes in information latency had no measurable effect in this study and small changes may not affect other studies as well. Technology effects might have to be considerable before there was any discernable difference in user performance.

The computer productivity paradox should also be reexamined after defining performance of groups studied to determine if it is the user performance traits that explain mixed results (Allnut, Haslam et al. 1990; Belardo and Pazer 1995; Goodhue and Thompson 1995; Triplett 1997; Brynjolfsson and Hitt 1998).

The large scale effect of user performance traits was not anticipated and future work should look at ways to discern differences between high performers and low performers in an organization but additionally, ways to upgrade the performance of lower achievers. Process changes brought about by latency avoidance should be studied. There may be other underlying reasons for less productivity that need to be discovered.

REFERENCES

Self Defense. Tennessee Code Annotated (TCA). **39-11-611**.

(1949). Brinegar v U.S., 338 US 160.

(1959). Draper v United States, 358 U.S. 307, 79 S.Ct 329. **3L. Ed.2d 327**.

(1985). United States v. Sharpe. **470 U.S. 675, 686**.

(1988). Criss v. City of Kent, 867 F.2d 259 (CA 6 Ohio).

(2000). Gardenhire v. Shubert, 205 F.3d 303 (CA 6 Tenn).

Agresti, A. and B. Finlay (1997). Statistical Methods for the Social Sciences. Upper Saddle River, NJ, Prentice Hall.

Allison, G. T. (1971). Essence of Decision: Explaining the Cuban Missile Crisis. Boston, MA, Little, Brown.

Allison, G. T. and P. Zelikow (1999). Essence of Decision: Explaining the Cuban Missile Crisis.

Allnut, M. F., M. H. Haslam, et al. (1990). Sustained Performance and Some Effects on the Design and Operation of Complex Systems. Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, London, England.

App., T. C. C. (2001). State v Levitt. 73 S.W.3d 159, Tenn. Crim. Ct. Appeals.

Banks, J., C. I. J, et al. (1996). Discrete-Event System Simulation. Upper Saddle River, N.J., Prentice Hall.

Beach, L. R. and R. Lipshitz (1993). Why classical theory is an inappropriate standard for evaluating and aiding most human decision making. Decision making in action: Models and Methods. G. A. Klein, J. Orasanu, R. Calderwood and C. Zsambok. Norwood, NJ, Ablex**:** 21-35.

Bechara, A., H. Damasio, et al. (1997). "Deciding advantageously before knowing the advantageous strategy." Science **275**(5304): 1293 - 1294.

Belardo, S. and H. L. Pazer (1995). "A Framework for Analyzing the Information Monitoring and Decision Support System Investment Tradeoff Dilemma: An Application to Crisis Management." IEEE Transactions on Engineering Management **42**(4): 352 - 359.

Brynjolfsson, E. and L. Hitt (1998). "Beyond the Productivity Paradox: Computers are the Catalyst for Bigger Changes." Communications of the ACM **41**(8): 49-55.

Brynjolfsson, E. and L. M. Hitt (2001). Beyond Computation: Information Technology, Organizational Transformation and Business Performance. The Puzzling Relations Between Computer and the Economy. N. Greenan, Y. Lhorty and J. Mairesse, MIT Press.

Carlson, J. R. and R. W. Zmud (1999). "Channel expansion theory and the experimental nature of media richness perceptions." Academy of Management Journal **42**(2): 153–170.

Cohen, M. S. (1993). The naturalistic basis of decision making. Decision making in action: Models and Methods. G. A. Klein, J. Orasanu, R. Calderwood and C. Zsambok. Norwood, NJ, Ablex**:** 205-227.

Cole, H., C. Vaught, et al. (1998). "Decision Making During a Simulated Mine Fire Escape." IEEE Transactions on Engineering Management **45**(2): 153 - 162.

Crandall, B. and K. Getchell-Reiter (1993). "Critical decision method: A technique for eliciting concrete assessment indicators from the "intuition" of NICU nurses." Advances in Nursing Sciences **16**(1): 42 - 51.

Daft, R. and R. Lengel (1986). "Organizational information requirements, media richness and structural design." Management Science **32**(5): 554–571.

Daft, R., R. Lengel, et al. (1987). "Message Equivocality, Media Selection, and Manager Performance: Implications on Information Systems." MIS Quarterly **11**(3): 355 - 366.

Daft, R., J. Sorumen, et al. (1988). "Chief Executive Scanning, Environmental Characteristics, and Company Performance: An Empirical Study." Strategic Management Journal **9**(2): 123 - 139.

Daft, R. L. and N. B. Macintosh (1981). "A Tentative Exploration into the Amount and Equivocality of Information Processing in Organizational Work Units." Administrative Science Quarterly **26**(2): 207-224.

de Groot, A. (1986). "Intuition in chess." International Computer Chess Association Journal **9**: 67-75.

De Silva, F. N. (2000). Challenges in Designing Spatial Decision Support Systems for Evacuation Planning. Aberdeen, Scotland, Department of Management Studies, University of Aberdeen, Scotland.

Dearing, B. (1990). "The Strategic Benefits of EDI." The Journal of Business Strategy **11**(1): 4-6.

Dennis, A. R. and S. T. Kinney (1998). "Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality." Information Systems Research **9**(3): 256–274.

Dennis, A. R. and J. S. Valacich (1999). "Rethinking Media Richness: Toward a Theory of Media Synchronicity." 1-10.

Drillings, M. and D. Serfaty (1997). Naturalistic Decision making in command and control. Naturalistic Decision Making. C. Zsambock and G. Klein. Mahwah, NJ, Lawrence Erlbaum Associates, Inc**:** 76.

Endsley, M. R. (1995). "Towards a theory of situation awareness." Human Factors **37**(1): 32-64.

Erev, I., G. Bornstein, et al. (1993). "The negative effect of probability assessments on decision quality." Organizational Behavior and Human Decision Processes **55**(1): 78 - 94.

Ericsson, K. A. and H. A. Simon (1993). Protocol analysis: verbal reports as data. Cambridge, MA, MIT Press.

Fischer, U., J. Orasanu, et al. (1995). Expert pilots' perceptions of problem situations. Eighth International Symposium on Aviation Psychology, Columbus, OH, Ohio State University Press.

Fulk, J. and B. Boyd (1991). "Emerging theories of communication in organizations." Journal of Management **17**(2): 407–446.

Gilliland, S., L. I. Benson, et al. (1998). "A Rejection Threshold in Justice Evaluations: Effects on Judgment and Decision-Making." Organizational Behavior and Human Decision Processes **76**(2): 113-131.

Goodhue, D. L. and R. L. Thompson (1995). "Task-Technology Fit and Individual Performance." MIS Quarterly **19**(2): 213 - 236.

Higa, K., O. R. L. Sheng, et al. (2000). "Understanding Relationships Among Teleworkers' E-Mail Usage, E-Mail Richness Perceptions, and E-Mail Productivity Perceptions Under a Software Engineering Environment." IEEE Transactions on Engineering Management **47**(2): 163-173.

Hoffman, R. R., B. W. Crandall, et al. (1998). "Use of the Critical Decision Method to Elicit Expert Knowledge: A Case Study in the Methodology of Cognitive Task Analysis." Human Factors **40**(2): 254 - 276.

Jamieson, R. (1996). Auditing and Electronic Commerce. EDI Forum, Perth, Western Australia.

Johansson, J. M., S. T. March, et al. (2003). "Modeling Network Latency and Parallel Processing in Distributed Database Design." Decision Sciences **34**(4): 676-706.

Johnson, E. J., J. W. Payne, et al. (1993). Adapting to Time Constraints. Time Pressure and Stress in Human Judgment and Decision Making. O. Svenson and A. J. Maule. New York, Plenum Press**:** 103-116.

Kaempf, G. L., G. Klein, et al. (1996). "Decision making in complex command-and-control environments." Human Factors **38**(2): 220-231.

Kahai, S. S. and R. B. Cooper (2003). "Exploring the Core Concepts of Media Richness Theory: The Impact of Cue Multiplicity and Feedback Immediacy on Decision Quality." Journal of Management Information Systems **20**(1): 263–299.

Kalakota, R. and A. B. Whinston (1996). Frontiers of Electronic Commerce, Addison-Wesley Publishing Company.

Kardash, C. A., J. M. Royere, et al. (1988). "Effects of schemata on both encoding and retrieval of information from prose." Journal of Educational Psychology **80**: 324-329.

Keller, K. L. and R. Staelin (1987). "Effects of Quality and Quantity of Information on Decision Effectiveness." Journal of Consumer Research **14**(2): 200-213.

Klein, G. (1998). Naturalistic Decision Making by experienced persons. Sources of Power: How People Make Decisions. Cambridge, Massachusetts, The MIT Press**:** 330.

Klein, G. (1998). Sources of Power: How People Make Decisions. Cambridge, MIT Press.

Klein, G., J. Orasanu, et al. (1993). Decision making in action: Models and methods. Norwood, NJ, Ablex.

Klein, G. A. (1989). Recognition primed decisions. Advances in man-machine systems research. W. B. Rouse. Greenwich, CT, JAI Press Inc. **5:** 47-92.

Lanir, Z. (1989). "The reasonable choice of disaster: The shooting down of the Libyan airliner on 21 February, 1973." Journal of Strategic Studies **12**: 479-493.

Lipshitz, R. and O. B. Shaul (1997). Schemata and mental models in recognition-primed decision making. Naturalistic Decision Making. C. Zsambok and G. Klein. Mahwah, NJ, Erlbaum**:** 293 - 304.

Maule, A. J., G. R. J. Hockey, et al. (2000). "Effects of Time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy." Acta Psychologica **104**: 283-301.

Miao, A. X., G. L. Zacharias, et al. (1997). "A Computational Situation Assessment Model for Nuclear Plant Operations." IEEE Transactions on Systems, Man and Cybernetics **27**(6): 728-742.

Orasanu, J. and T. Connolly (1993). The Reinvention of Decision Making. Decision Making In Action: Models and Methods. G. Klein, J. Orasanu, R. Calderwood and C. E. Zsambok. Norwood, N.J., Ablex**:** 3 - 20.

Ordonez, L. and L. I. Benson (1997). "Decisions under Time Pressure: How Time Constraint Affects Risky Decision Making." Organizational Behavior and Human Decision Processes **71**(2): 121-140.

Pascual, R. and S. Henderson (1997). Evidence of Naturalistic Decision Making in Military Command and Control. Naturalistic Decision Making. C. Zsambock and G. Klein. Mahwah, NJ, Lawrence Erlbaum Associates, Inc**:** 217-226.

Pennington, N. and R. Hastie (1993). "Reasoning in explanation-based decision making." Cognition **49**: 123 - 163.

Raiman, O. (1990). A Circumscribed Diagnosis Engine. Proceedings of the International Workshop on Expert Systems in Engineering, Principles and Applications, London, UK, Springer-Verlag.

Rice, R. E. (1992). "Task Analyzability, Use of New Media, and Effectiveness: A Multi-site Exploration of Media Richness." Organizational Science **3**(4): 475-500.

Rice, R. E. (1993). "Media appropriateness: Using social presence theory to compare traditional and new organizational media." Human Communication Research **19**(4): 451–484.

Roberts, N. C. and J. D. Dotterway (1995). "The Vincennes Incident: Another Player on the Stage?" Defense Analysis **11**(1): 31-45.

Roper, W. (2002). Information Systems for Infrastructure Management and Emergency Response. 34th Joint Meeting Panel on Wind and Seismic Effects, Japan.

Ross, J. and B. M. Staw (1993). "Organizational Escalation and Exit: Lessons from the Shoreham Nuclear Power Plant." Academy of Management Journal **36**(4): 701 - 732.

Schaafstal, A. M., J. H. Johnston, et al. (2001). "Training teams for emergency management." Computers in Human Behavior **17**: 615-626.

Schmitt, J. F. and G. Klein (1996). Fighting in the fog: Dealing with battlefield uncertainty. Marine Corps Gazette**:** 62 - 69.

Shadish, W. R., T. D. Cook, et al. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston, Houghton-Mifflin.

Shanteau, J. (1992). "Competence in Experts: The role of task characteristics." Organizational Behavior and Human Decision Processes **53**: 252-266.

Simon, H. A. (1957). Models of man: social and rational. New York, Wiley.

Srinivasan, A., D. Sundaram, et al. (2000). Implementing Decision Support Systems: Methods, Techniques and Tools. Maidenhead, U.K., McGraw-Hill.

Svenson, O. and L. I. Benson (1993). Framing and Time Pressure in Decision Making. Time Pressure and Stress in Human Judgment and Decision Making. O. Svenson and A. J. Maule. New York, Plenum Press**:** 133-144.

Swets, J. A. (1992). "The Science of Choosing the Right Decision Threshold in High Stakes Diagnostics." American Psychologist **47**(4): 522-532.

Triplett, J. E. (1997). The SOLOW Productivity Paradox: What Do Computers Do to Productivity? Conference on Service Sector Productivity and the Productivity Paradox, Ottawa, Canada, Brookings Institution.

Tversky, A. and D. Kahneman (1974). "Judgment under uncertainty: Heuristic and biases." Science **185**(4157): 1124 - 1131.

Weaver, E. and G. Richardson (2002). Threshold Setting and the Cycling of a Decision Threshold. 20th Systems Dynamics Conference, Palermo, Italy.

Weick, K. (1993). "The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster." Administrative Science Quarterly **38**(4): 628-652.

Wright, P. (1974). "The harassed decision maker: Time Pressure, distractions and the use of evidence." Journal of Applied Psychology **59**(5): 555-561.

Zsambok, C. E. (1997). Naturalistic Decision Making: Where are we now? Naturalistic Decision Making. C. E. Zsambok and G. Klein. Mahwah, NJ, Lawrence Erlbaum Associates, Inc.