

**COURSE OUTLINE - PSYCHOLOGY 446****EYEWITNESS PSYCHOLOGY****Dr. R. LINDSAY**

Fall term 2013

Contacting the instructor: I am terrible at responding to phone calls. I have been known to ignore voice mail messages for weeks at a time! I check my email every day that I am in. As a result, it is much better to email me than to phone. So...

Email: **rod.lindsay@queensu.ca**

Phone: 533-2880

**Readings:** The required readings for the course consist of chapters from the **Handbook of Eyewitness Psychology**.

**Evaluation:** Course grades are obtained from four sources:

- 1) 30% Seminar presentation and leading the discussion.
- 2) 20% class participation (10% attendance, 10% being active).
- 3) 10% thought papers (1 thought paper per class after the seminars begin in week 4)
- 4) 40% research proposal (details below).

The participation, seminar, and research proposal will be marked via letter grades then combined at the end of the course. Attendance and thought papers will be “scored” as % of thought papers submitted and classes attended. There is no reason generally to miss a thought paper given that they are submitted via email. If you must miss classes it will make no difference until you are missing more than 5 of the classes with seminar presentations. If you must miss so many classes

**Evaluation in Detail:**

**NOTE: All assignments are to be submitted electronically and using PDF format only!**

**All grades will be letter grades so perfect attendance (for example) would be an A, not 100%.**

1) **Thought Papers:** Starting on the first class of the fourth week of the course, specific topics will be covered with in-class presentations. There are assigned readings for each class from that date on. For each of the classes on and after that date, thought papers are submitted. A thought paper is literally that – read the assigned readings and write a brief paper (a single page is sufficient, never more than 2) outlining your thoughts about the research and issues discussed. Note that all thought papers will be submitted and shared with the entire class via email.

2) **Seminar presentations:** Students (generally in pairs) are responsible for a seminar presentation. The class will have read the assigned readings. The purpose of the seminars is to *briefly* describe the issues

and general trends in the area, *bring the topic to life* (via demonstrations, videos, etc.) and *encourage discussion*. You will 70 of the 80 minutes of class time for your presentations. The remaining 10 minutes I reserve for me to add comments, interpretations, anecdotes, etc though I may not always do so.

3) **Class participation:** Class participation is just that. In classes such as this one, people often attend less and less after they present their seminars. This is discourteous to their classmates and eventually erodes discussion as fewer people are available to contribute. Attendance will be taken at each class (you will not lose marks for missing a few classes but overall poor attendance will cost you). Attendance alone will not ensure all of the marks; you also need to take part in class discussions, bring up ideas, etc. Perfect attendance will guarantee 10 marks, after that, marks are earned by contributions. On the other hand, you should not feel that you must comment on every issue.

Note: a) for the first few classes at least I would like students to identify themselves when they ask or answer questions so that we all get to know each other. Also, I shall ask specific students to comment from time to time, particularly those who do not voluntarily do so.

4) **Research proposal:** The research proposal is worth 40% of the course grade and is due via email no later than the end of the last week of classes (yes midnight Friday will do). The format is quite short. It is based on the grants that I write for my studies. We will use the SSHRC format provided below. Note that it is only a 6 page document! PLEASE NOTE: even though you may be sharing a presentation, your research proposals must be completely independent – no collaborative projects are permitted.

The seminar schedule will be emailed separately (possibly already has been) and contains the assignments for readings as well.

The next few pages provide the following:

1. The current SSHRC outline for preparing a grant application (with some comments about sections not appropriate to you).
2. Two samples of applications.

SSHRC outline and two examples of actual SSHRC grant applications. Note that these must be submitted in size 12 font and as PDF files.

**Section 1. A title followed by a one page summary of the proposed project. You will not provide a summary so ignore this part of the SSHRC application.**

**Section 2. The next section is broken down into multiple subsections and you are restricted to 6 pages totally for all of it! This is your paper!**

## **2. Title**

### **A. OBJECTIVES**

- 1) **Short-Term Objectives:** What is the question that the data to be collected directly addresses?
- 2) **Long-Term Objectives:** In the long term, how could this knowledge be applied or useful?

### **B. CONTEXT**

- (1) **Relevant Scholarly Literature:** Briefly review the relevant literature.
- (2) **Rationale of the Proposed Program of Research:** Explain the logic of the study.
- (3) **Relationship of Proposed Program of Research to Previous SSHRCC Funded Research:** (just put the heading followed by “New scholar” which will give you a bit more room than I typically have).
- (4) **Contributions to knowledge:** How will the results of the proposed study increase our knowledge of eyewitness (or related) issues?

**C. METHODOLOGY:** Describe in as much detail as you can fit in how the data will be collected, relating it clearly to the issues to be tested. **(New scholars usually write more detailed methods because they have not established that they will be successful researchers on their own while experienced researchers have a bit more wiggle room to be imprecise because their publication record is used as an important factor in evaluating their applications.)**

**D. COMMUNICATION OF RESULTS:** Sufficient to say results would be submitted to conferences and for publication in journals.

## Sample SSHRC applications

**2. DETAILED DESCRIPTION****Radical Lineups = Fewer Wrongful Convictions****A. OBJECTIVES**

**(1) Short term objectives.** To test a radical new lineup procedure designed to reduce false positive choices and calibrate the accuracy of the suspect choices from lineups.

**(2) Long term objectives.** To reduce the rate of wrongful convictions based on mistaken eyewitness evidence from lineups and provide a procedure that works with all witnesses regardless of race, age, etc.

**B. CONTEXT**

**(1) Relevant Scholarly Literature.** Despite some success at reducing the high rate of false positive choices from lineups (Lindsay & Wells, 1985), eyewitness errors remain the major cause of wrongful conviction (Connor et al, 1996). This year alone, mistaken selections from lineups will lead to the wrongful conviction of at least 40 and perhaps 200 to 300 Canadians (Lindsay, 2003). We need an identification technique that 1) dramatically reduces false positive choices, 2) does not dramatically reduce correct selections, and 3) permits an evaluation of the likely accuracy of eyewitnesses based on their lineup decisions. I propose to study the effectiveness of multiple, large, independent and quasi-independent lineups as a means of generating eyewitness evidence that is easy to obtain, highly diagnostic of guilt and innocence, and has the potential to provide calibrated evidence of the likelihood of guilt (Pryke, Lindsay, Dysart, & Dupuis, 2004).

Lindsay and Wells (1985) produced the first successful modification of lineup procedures. They demonstrated that presenting lineup members individually and requiring a decision (the person is or is not the criminal) resulted in a small decline in correct selections when the criminal was present but a large decline in false choices when the criminal was absent from the lineup. Subsequent studies confirmed that the result was replicable and that the "sequential" lineup reduced false positive choices in general (Stebly et al, 2001) as well as when lineups were biased by the use of poor foils or instructions (Lindsay, Lea, Nosworthy, Fulford, Hector, LeVan, & Seabrook, 1991).

Other researchers attempted to improve eyewitness accuracy by providing multiple cues via videotaped or live lineups. The logic was that cues such as voice and motion would assist the witness so that the rate of inaccurate choices made from photo arrays would be reduced. Cutler, Berman, Penrod, and Fisher (1994) conducted a meta-analysis of the literature and concluded that additional cues presented together in lineups produced a trivial effect on accuracy, possibly because one cue, the face, overwhelms the other features. Witnesses select lineup members based on the face and then convince themselves that inconsistencies of body or voice are not sufficient to alter their decision.

Levi (1998) attempted to resolve this problem by increasing the size of the lineup. Using 10-, 20-, and 40-person, videotaped, sequential lineups he was able to demonstrate low rates of false choices of specific individuals. However, the overall rate of wrong choices increased linearly as lineup size increased. As a result, the expected rate of incorrect selection of any randomly selected lineup member was not changed by an increase in lineup size alone. Furthermore, the rate of correct rejection (selecting no one from a criminal-absent lineup) was 65% compared to 72% with the original, 6-person, sequential lineup (Stebly et al, 2001). As

a result, Levi's approach produced no improvement in correct decisions when the target is absent from the lineup (as compared to the Lindsay and Wells 1985 procedure).

**(2) Rationale of the Proposed Program of Research.** If the features were presented independently, then no feature could overwhelm another. Witnesses in the proposed studies will have multiple opportunities to select various parts of the person including the face, body, and voice. Multiple independent selections of various features of the same person increase the likelihood of that person being the one who was previously encountered by the witness. Conversely, inability to select someone from more than a single lineup should indicate a lower probability of guilt.

Selection from independent lineups ought to be associated with very low rates of multiple false choices (choosing the same innocent individual from multiple, independent lineups) since the innocent suspect will rarely resemble the guilty party more than all other lineup members on all features tested. If the lineups are independent, the probability of multiple false choices can be calculated based solely on the size and number of lineups employed.

A multiple independent lineup procedure also will "calibrate" the likely accuracy of eyewitnesses. Combining independent selections of features will produce evidence that can assist in estimating if the suspect is guilty. The more often someone is selected from independent lineups, the more likely it is that the person is the criminal. Selection of an innocent suspect from 2, independent, 10-person lineups should occur only .01 of the time (because the suspect is innocent, the witness must guess, and selection should occur only at a chance level from each lineup). Selection from 3 such lineups would occur at a rate of .001, from four at a rate of .0001, etc. The frequency of multiple correct selections is an empirical issue to be established by research. The ratio of these two rates (the diagnosticity ratio, Wells & Lindsay, 1980) is a measure of the probative value of the resulting evidence.

**(3) Relationship of Proposed Program of Research to Previous SSHRCC Funded Research.** Pryke et al (in press) used face, body, and voice lineups to produce evidence with low odds of being associated with innocent suspects (1/216) while being common for guilty suspects (1/3). Selection from 2 or more lineups was rare for innocent people (3/100) but frequent for the guilty party (1/2). Selection from 3 lineups occurred infrequently for targets (about 1/10) but never for anyone else. Boyce, Lindsay, and Dupuis (2003) replicated and extended this finding using rankings rather than selection of lineup members. Some patterns of response were associated only with accurate selection and others only with inaccurate selection. The proposed studies will follow from this work by increasing both the number of lineups (by including profile lineups) and the size of lineups (from 6 to at least 20). Traditional, simultaneous lineups as large as 20 have been shown to produce comparable correct selection rates to 6-person lineups (Nosworthy & Lindsay, 1990).

**(4) Contributions to knowledge.** Current practices will generate a rate of from 1% to 9% choices of innocent suspects when the criminal is not in the lineup depending on the procedure used and presence or absence of lineup biases (Lindsay, 2003; Steblay et al., 2001). The proposed methods will result in a rate of no more than 1 chance in 2400 of obtaining such a dangerous result ( $1/(20^2 \times 6)$ ). There is a remote possibility that the rate could fall to nearly one in a million ( $1/(20^4 \times 6)$ ).

**C. Methodology.** People will view "target" individuals then be asked to attempt to select the targets from lineups of full faces, face profiles, full bodies, body profiles, and voices. The photographic lineups will contain at least 20 individuals. The voice lineup will contain about

6 (pilot work suggests that large voice lineups are impractical). The probability of choosing the same innocent person by chance from all five lineups is  $1/(20^4 \times 6)$  or 1/960,000 assuming that every witness chooses from every lineup, the lineups are fair, and the lineups are truly independent.

**Non-choosing.** Given that witnesses are not required to choose anyone, the true probability may be much lower. If only 50% of witnesses select from each lineup, the probability of 5 coincidental wrong choices would be approximately 1 out of 31 million, roughly the population of Canada!

However, the lineups also must be "fair" and "independent" for these calculations to be valid. Failing to meet these criteria would dramatically alter, and almost certainly increase, the probability of the same innocent person being selected from multiple lineups. The research will involve testing how difficult it is to achieve fairness and independence.

**Lineup fairness.** A lineup is fair if, on average, the suspect is no more likely to be selected than other lineup members by people who have not seen him but are aware of his description (Malpass & Lindsay, 1999). An advantage of the multiple lineup technique is that overall bias may be difficult to produce because the same innocent person is unlikely to stand out across multiple features given that the lineups are fair. Thus, an innocent suspect that stands out in a standard lineup because his face is similar to the true criminal may have a voice and body not easily mistaken for the criminal's. Descriptions of bodies and voices tend to be minimal and vague making the production of fair lineups relatively easy.

Lineup fairness measures using the "mock witness" procedure will be collected. The results will be used to calculate modified expectations of individual and multiple incorrect selections to be compared with the actual selection data. Proportion of choices from the mock witness task has been shown to "postdict" such errors with witnesses shown individual lineups and thus may provide important information for evaluating the fairness of multiple lineups as well (Lindsay, Smith, & Pryke, 1999).

**Lineup independence.** Independence of full face, full body, and voice lineups is easily achieved (Pryke et al., in press). Selections of full and profile faces and full and profile bodies likely will not be completely independent. However, even if the profiles were abandoned, multiple identification of an innocent person would be expected no more than 1 in 2400 times [ $1/(20^2 \times 6)$ ]; and less if not all witnesses select from all lineups. Partial dependence still may result in a useful contribution from the additional lineups (see next point).

**Partial selection.** The rate of partial selection (from some but not all lineups) of both guilty and innocent suspects is important. The multiple, independent, large lineup approach will be useful only to the extent that people are able to select previously seen targets from multiple lineups of different features. The more lineups used, the more chances there are for multiple selections of both guilty and innocent suspects; thus the addition of profile lineups. It is an empirical issue whether the addition of profile lineups that are not completely independent will improve the discrimination between lineups containing guilty and innocent suspects.

**False positives.** The method can not fail to generate low, multiple, false-positive selection rates so long as fairness and independence are achieved. If all lineup members fit the description of the relevant features provided by the witness, and the suspect is innocent, then the witness must guess and the laws of probability inform us of the likely rate of the various possible patterns of choice.

**Correct selection.** Rates of multiple correct selection will be determined empirically. The rate of multiple selections of the previously seen person must be high enough to justify adopting the procedure. Pryke et al (in press) found that about 65% of witnesses who selected the face of the guilty party also were able to select him from at least one of their two additional, independent lineups; a promising start.

**Confidence, rankings, and ratings.** Another possibility is that even more diagnostic data can be obtained using confidence judgments, rankings, or ratings of lineup members in addition to or instead of selection decisions. Selection provides only a binary value. Selections could be weighted by confidence in those selections to see if the combined selection and confidence data from multiple lineups is highly diagnostic of guilt. Alternatively, witnesses asked to rank or rate all lineup members in terms of their similarity to the witnesses' memory of the criminal may generate data that are highly predictive of guilt and innocence. Dupuis and Lindsay (2001) found that summing the ranks assigned to suspects from 6-person face, body, and voice lineups produced a scale that contained values at one extreme that were completely predictive of guilt and at the other extreme were completely predictive of innocence.

**Staged Crime Studies.** To increase the probability that the results generalize, participants will view one of multiple targets (at least 14, possibly as many as 50). Following exposure to the target, selection procedures will be conducted. Data will be collected using standard, 6- and 12-person, simultaneous and sequential, target-present and target-absent lineups for each target to provide comparison data with current procedures.

Multiple, large, independent lineups will provide participants with an opportunity to select the target's face, body, voice, body profile, and face profile. The profile lineups will always be shown last (at least in initial studies) because they are unlikely to be completely independent of the face and body non-profile lineups. By presenting them last, the data from the first three lineups can always be analyzed as if the others had not been presented.

All of the "techniques" previously mentioned will be tested. In some studies, participants will attempt to select the target and rate their confidence. In other studies, participants will rate or rank all lineup members from most to least similar to their memory of the target.

Data analyses will start with the construction of scales by summing the results of the decisions. Thus, the target will be associated with a number from 0 to 5 to reflect the number of times the target was selected from the 5 lineups, or that value weighted by the confidence of those choices, or the sum of his rankings or ratings. Data analysis for target-absent conditions will follow a "worst case scenario" approach. The individual who would, across the entire study, generate scale values most associated with guilt will be treated as the innocent suspect to provide a stringent test of the ability of the technique to protect innocent suspects. Other analyses will use random assignment to the status of innocent suspect to estimate more typical results.

These studies will require very large sample sizes because we are attempting to demonstrate a significant decline in false positive choices from the (at most) 9% rate obtainable with current procedures. A "floor effect" requires considerable power to demonstrate success. The plan is to run as many witnesses as possible in these studies over the term of the grant. My goal is to collect data from about 6000 such participants during the 3 year period.

**Independence Studies.** To test the independence of the lineups, other participants will be exposed to a member of a lineup (e.g., face) and be asked to select the same person from

another lineup (e.g., body). By following this procedure for all lineup members and all lineups, it will be possible to determine if the various features can be matched. Completely chance patterns in the results would indicate true independence. Deviation from chance indicating that the various features can be associated with each other without exposure to the "full" person indicates that the lineups are not completely independent. Achieving independence has the benefit of permitting calculations of probabilities as discussed earlier. However, failure to achieve complete independence does not mean that the multiple lineup technique is not useable. In the absence of complete independence, we must rely on the empirical data to indicate likely false selection rates. These rates may still be substantially below current rates. One hypothesis to be tested will be the possibility that the degree of independence can be calculated from lineup independence tasks and then used to estimate likely false positive selection rates. This would parallel the current use of lineup fairness data to estimate false positive selections from individual lineups (Smith et al, 1999). Independence studies will require fewer participants, perhaps 300 totally.

**Lineup Fairness Studies.** Measures of lineup fairness are based on the "mock-witness" task (Malpass & Lindsay, 1999). Using the descriptions provided by witnesses to the staged crimes, others (mock-witnesses) select lineup members based solely on the descriptions provided by the actual witnesses. The proportion of target choices from this task provides a measure of lineup fairness and is highly predictive of selection of innocent suspects from target-absent face lineups (Lindsay, Smith, & Pryke, 1999). Fairness studies will require several hundred participants.

**Additional Studies.** Once these studies have established (as I believe they will) that multiple independent lineups are superior to currently used procedures, it will be necessary to address a number of traditional issues in the eyewitness area. I will need to explore the degree to which the results generalize across race of target and witness (cross-race effects) and age of witness (children and elderly witnesses). Field studies will be needed to ensure that any effects are not restricted to undergraduates as participants. The effects of disguise and multiple perpetrators on the effectiveness of the procedure will need to be explored. These studies require approximately 1500 participants. The scope of this proposal probably will require more than 3 years to complete. I simply plan to start and get as far as I can in 3 years.

#### **D. COMMUNICATION OF RESULTS**

Research results will be submitted to peer reviewed journals (e.g., Applied Cognitive Psychology, Law and Human Behavior) and presented at professional meetings of psychologists (e.g., American Psychology-Law Society, Society for Applied Research in Memory And Cognition). Presentations also will be made to professional groups such as police, lawyers, and judges through conferences and workshops. Graduate students will be encouraged to present at least one paper each year at a relevant conference. I normally present several papers each year, including invited addresses.

Comments for 446 students. You do not have previous SSHRC supported research and should provide a comment that you are a new scholar and use the space for other purposes. I provided only a very general description of the work to be done. I can get away with this because I am an established scholar. Based on prior publication in the field they accept that I can design and conduct studies that pass the test of peer review. As a new scholar, the review panel would be much more concerned with the precise details of your planned research. The flip side of this is that they expect people like me to propose grandiose things such as this (multiple studies) while new scholars are only expected to describe something useful. Finally, you will notice that although references were provided in text, there was no reference section at the end of the paper. Actually the references are provided but are not counted in the 6 pages so you too should include references but start a new, seventh page for those. Note that this sample is less than 6 full pages. Shorter is acceptable, longer is not.

Another example follows:

**Detailed Description: Pattern jury instructions re Proof Beyond a Reasonable Doubt (PBRD)**

**A. OBJECTIVES**

**Short term objectives:** We will examine the use of the PBRD standard, specifically studying:

- (1) Understanding: How do legal systems, lay people, and justice system professionals understand and use proof beyond a reasonable doubt?
- (2) Current phrasing: What is the impact of the current Canadian instructions concerning PBRD (*R v Lifchus*, 1997) on the assessment of the credibility and reliability of child and adult witnesses?
- (3) Impact of errors: What is the effect of mis-stating or omitting portions of the current articulation of PBRD instructions on the assessments of the witness credibility and reliability?

**Long-Term Objectives:** To propose reforms to law and practice that are empirically demonstrated to have psychological validity for explaining the concept of PBRD to jurors and justice professionals, and to increase psychological knowledge about the assessment witness credibility and reliability.

**B. CONTEXT**

**1) Proof Beyond a Reasonable Doubt:** In criminal trials, it is usually impossible to determine with absolute certainty what actually occurred. The concept of PBRD has long been central to the criminal justice system, guiding decision-making in the face of uncertainty (DeLoggio, 1986; Shapiro, 1991), balancing an accused's right to freedom and society's right to be protected from crime (Hamer, 2004). PBRD is the highest standard of proof, and is an instrument for reducing the risk of wrongful convictions (*In re Winship*, 1970; Dhami, 2008). However, the higher the standard that the prosecution must meet, the greater the risk of unjustified acquittal (Arks & Malloy, 2002; Hamer, 2004).

Reasonable doubt is often the pivotal issue in a criminal trial, with the defense arguing that the evidence is insufficient to reach this standard of proof. As a result, legal systems have frequently addressed the issue of what reasonable doubt means and how best to communicate that meaning to jurors (Hemmens, Scarborough, & Del Carmen, 1997). However, reasonable doubt is a difficult concept for legal experts to define and even more difficult to explain to jurors who are untrained in legal terminology and concepts (Wright & Hall, 1997).

There is debate whether judges should even try to define the PBRD concept (*Gaines vs. Kelly*, 2000), and there is some empirical support for not doing so. In a preliminary study, one of our students, Nugent (2008), found a decrease in laypersons belief of children's reports of events from 58%, without mention of PBRD, to 36% when asked to consider PBRD, even without providing any definition. This suggests that merely mentioning the PBRD standard may be sufficient to increase the proof people demand. However, it is generally accepted that it is better that a judge define the phrase rather than leave it open to interpretation to jurors (Horowitz, 1997). These instructions can help to clarify the concept but can also increase confusion (Kerr et al, 1976). To explain PBRD, American case law commonly uses terms like: "doubt based on reason", "serious and substantial doubt", "doubt that can be articulated", a reason to "hesitate to act in important affairs of life" (Hemmes et al, 1997). Alternatively explanations such as "having an abiding conviction" or "a moral certitude" are suggested as thresholds above which the juror can feel comfortable rendering a guilty verdict based on absence of reasonable doubts (Horwitz, 1992). Stoffelmayr and Diamond (2000) propose that four criteria are central to the legal and empirical considerations of reasonable doubt instructions: "(a) absolute certainty should not be required, (b) a high threshold for conviction should be specified, (c) the beyond a reasonable doubt standard should be distinguishable from lower standards of proof, and (d) the instruction should encourage consistent application by jurors sitting on the same case" (pg 770).

To produce consistency in decision making, many jurisdictions have developed recommended instructions ("*pattern jury instructions*") that the judge provides to a jury, and that judges are to follow in

their own decision-making if there is no jury. In Canada, trial judges are required to instruct juries about the concept in their charge to the jury, and to demonstrate a correct understanding of it in cases that are decided without a jury. Trial decisions may be reversed if the judge has failed to follow an accepted pattern jury instruction (*R v Lifchus*, 1997), as the failure to follow the recommended phrasing may have resulted in decisions made using an inappropriate standard of proof (Ellsworth & Reifman, 2000; Finkel, 2000). Appeals based on errors when instructing the jury are common; a significant portion of these appeals result in orders for new trials, though in practice the prosecution is often discontinued after an appeal, especially if the case involves a child witness. In Canada, crown prosecutors must also make initial PBRD determinations, and only proceed with a prosecution if satisfied that there is a “reasonable prospect of conviction” on the PBRD standard. Variability in the applicability of the PBRD standard by prosecutors can lead to unfairness in the legal process.

**2) Psychological Research:** In general, research (largely American) on the degree to which lay persons understand pattern jury instructions indicates poor levels of comprehension (e.g., Charrow & Charrow, 1979; Elwork, Sales, & Alfini, 1982; Strawn & Buchanan, 1976). In one study, mock jurors did only 6% better in answering questions about the law after they heard Washington State pattern jury instructions, a result of particular interest because reasonable doubt instructions were included in this study and hearing the instructions had no effect on understanding the concept of reasonable doubt (Severance & Loftus, 1982). Survey studies have found that without an instruction, many potential jurors wrongly believe reasonable doubt is the same as absolute certainty (Kramer & Koenig, 1990; Montgomery, 1998), and many legal professionals also hold this view (Zander, 2000). Some suggest that the standard should represent a value above 90% probability of guilt (Newman, 1993). However, studies have found that jurors exposed to this standard of PBRD can make guilty judgments with between 60% and 90% subjective certainty in guilt (Hashe, 1993; Dhimi 2008). Having only a 60% belief in guilt does not seem appropriate for the PBRD standard. Psychological research suggests that defining reasonable doubt in terms of probabilities can improve decision making (Kagehiro 1990). Kagehiro and Stanton (1985) found quantitative instructions reduced guilty verdicts, whereas qualitative instructions did not (conviction rates were similar regardless of the jury instructions used). However, judges and lawyers often object to quantitative instructions (Note, 1995; *R. v. McLeod*, 2010).

There is a need to understand how PBRD instructions affect decision making of jurors and justice system professionals. The PBRD standard is, in legal theory, more demanding than the civil standard of proof on the balance of probabilities. It should raise the threshold of belief in guilt in order to obtain a conviction, by inviting jurors to entertain the possibility that the defendant is legally innocent even if they believe that he in fact committed the act in question. However, in an American study, Wright and Hall (2007) found that PBRD instructions caused some lay participants to lower their threshold of belief, and used it to justify guilty verdicts. Horowitz and Kirkpatrick (1996) hypothesized that if reasonable doubt instructions are effective, properly instructed people asked to decide cases should be less willing to convict based on weaker evidence. In their study of mock jurors it was found that if they were provided no definition or one that only expressed the idea that PBRD requires evidence “which is stable and does not vacillate”, they did not distinguish in their assessments of guilt between two cases, one with strong, and the other with weak, evidence of guilt. If provided with a definition emphasizing the need for real doubt, the mock jurors were much more likely to make a distinction in guilt findings between cases with strong and weak evidence, but this study did not find a difference in outcomes between use of two short definitions, emphasizing that it must be a “real doubt” and a full version of a court approved pattern jury instruction (*Victor v Nebraska*, 1994). This American study suggests that it is important to offer clear guidance to jurors about the meaning of PBRD, but raises doubts about the significance of the exact words used.

**3) Canadian Legal Context:** Canadian appeal courts are increasingly encouraging the use of pattern jury instructions (Comisky, 2010), but judges must also be prepared to answer queries from juries about the meaning of the concept that go beyond restating the pattern instructions (*R v Layton*, 2009). There has been

very little research on the comprehensibility and effect of different articulations of PBRD, and none about the standard presently used in Canada, and its effects on the assessment of the credibility of witnesses. Canadian jury instructions tend to be longer than in the USA, as the judge's charge to the jury includes more information (a summary of the evidence). As a result, research conducted in the United States may not generalize to Canada where almost no research on the topic has been published. For this reason, Comisky (2010) recently pointed out that "systematic studies are required to test overall comprehensibility of the Canadian [pattern jury] instructions" (pg. 648). Rose and Ogloff (2001) conducted the only reported study of the comprehensibility of Canadian pattern jury instructions, focusing exclusively on instructions about conspiracy law. They concluded, based on responses to true-false questions about understanding of mock jurors, that hearing the standardized instruction did not increase understanding of this complex area of law.

**4) Credibility Assessments:** Reasonable doubt instructions are intended to create a sufficiently strict standard of proof that accused persons are not convicted based on false or weak evidence, but at the same time a standard that allows for the successful prosecution of those who are clearly guilty. Cases with weak evidence should be less likely to result in conviction than cases with strong evidence; this outcome should be accentuated by PBRD instructions, such that the more effective the instructions, the lower the conviction rate should be with weak evidence, while cases with strong evidence should be less influenced or not influenced at all if the evidence is strong enough. Evidence evaluators (e.g., lawyers, judges, jurors) have to assess the credibility and reliability of testimony. The influence of PBRD instructions on an individual evaluator may vary depending on a number of factors: (a) *Age of witness*: Research suggests that judgments of child witnesses' credibility are often based on general beliefs and stereotypes about children (e.g., Goodman, et al, 1989; Haugaard & Reppucci, 1992; Nikonova & Ogloff, 2001). Compared to adults, child witnesses are perceived as less cognitively competent and more susceptible to suggestion, but more trustworthy and sincere (Brigham & Spier, 1992; Goodman et al., 1989; Ross et al., 1990; Yarmey & Jones, 1983). (b) *Type of Event*: Perceptions of children's reports may differ according to whether they are victims (e.g., in abuse/assault) or bystanders to a witnessed crime like a theft (Gabora, Spanos & Jacob, 1993; Goodman et al., 1998). (c) *Individual Characteristics of Evaluators*: The characteristics of those undertaking a credibility assessment may influence the judgment of witness reports (Porter, Campbell, & Stapleton, 2002). For instance, women jurors are generally more likely to believe prosecution witnesses than men (Wiley & Bottoms, 2009). (d) *Eyewitness identification*: Police lineup identification evidence is notoriously unreliable in a wide variety of circumstances (Lindsay & Pozzulo, 1999) and people have difficulty discriminating accurate from inaccurate witnesses (Boyce, Beaudry, & Lindsay, 2007). A stringent standard of proof is recommended in cases where identification is the only or central evidence (Cory, 2001). (e) *Laypersons vs. Professionals*: Some researchers suggested that professionals with extensive justice system experience may be better at credibility assessment (O'Sullivan, 2005; Vrij & Mann, 2005). In some studies law enforcement professionals had high detection rates of adults' lies (Ekman & O'Sullivan, 1991; O'Sullivan, 2005), but they were at chance levels in others (e.g., Porter et al, 2000; Vrij, Edward & Bull, 2001). They often do not fare any better than laypersons (Leach et al 2004; Leach et al, 2009). It is an open question whether credibility assessment by justice system professionals based on a PBRD standard may differ from those of laypersons. Further, no existing studies have systematically examined the impact of the factors listed above on PBRD assessments; we will be address this gap in knowledge.

Other important, previously unaddressed issues will be studied. Among them are *instruction errors*: Pattern jury instructions regarding reasonable doubt are complex rather than a single, simple statement. Appeal courts may order a new trial based on the failure to provide a complete instruction, based on the assumption that each part is critical to convey the meaning of reasonable doubt. To test this, we will break down the present Canadian PBRD instructions into subsections and test to determine the degree to which the various subsections significantly alter understanding of the concept of PBRD, credibility assessments, and findings of guilt. This work has great potential value as it may be that omitting some portions of the

instructions produces little or no difference in understanding, while the omission of other portions is critical. Knowing which sections are critical could influence judicial practice and appellate jurisprudence.

**5) Rationale of the Proposed Program of Research:** Our program of research will address the significant gaps in the literature outlined above by examining the influence of PBRD instructions on credibility assessments of child and adult witnesses by justice system professionals and potential jurors. Our multidisciplinary perspective is essential to the success of the project. An interdisciplinary team can integrate the legal approach, based on an adversarial model of justice, with its legitimate concerns about the rights of the accused and fair process, with a psychological approach, with its use of experimental manipulations (e.g., counter-balancing, randomization) to rule out confounding factors. Further, given the scarcity of research on PBRD, especially in the Canadian context, *and* the lack of forensically relevant research, it is important to learn both what justice system professionals believe about the utility of PBRD instructions and their use of the principle of PBRD for assessing cases compared to laypersons' interpretations of PBRD. Most importantly, a multidisciplinary program of research bridges the gap in the literatures of both law and psychology. To date, there are serious limits both in the understanding of the effect of the PBRD standard in the justice system (e.g. lack of information on the methods and criteria used by justice system professionals for assessing PBRD), and in the psychological literature (limited research on the influence of PBRD instructions on layperson's judgments). The proposed program of research involves the multiple actors in the legal process: laypersons (potential jurors), lawyers, and judges.

**6) Relationship of Proposed Research to our Previous Research:** The proposed research significantly builds on our multiple SSHRC-funded (1999-2011) interdisciplinary research agenda. The findings from these research grants have led to significant reforms in the Canadian criminal law and provide strong empirical evidence for judicial practice and education (see below). Nugent's (2008) thesis was conducted as part of this research and directly led to our current interest in PBRD. The proposed program of research builds upon and significantly extends our existing work to examine important issues related to the administration of justice and forensic psychology. This research will: 1) build on our research examining factors that influence adult credibility assessments of child witnesses; 2) utilize our ability, based on previous collaborations, to generate true and false reports (using both ethically and ecologically valid methods) of criminal acts that allow for the investigation of adults' abilities to accurately judge which are true (guilty verdict appropriate) and which are false (not guilty verdict appropriate); 3) build on our expertise to generate correct and erroneous eyewitness identifications; 4) build on our expertise surveying and interviewing justice system professionals regarding attitudes and evaluations of current legal practices; 5) employ research paradigms that we have expertise using, namely, examining the veracity of reports of assault and the accuracy of eyewitness identification.

**7) Contributions to knowledge:** Our SSHRC funded research has made significant contributions to both psychological and legal knowledge about credibility assessment of child and adult witnesses, and has resulted in significant improvements in how children's competence is assessed in Canadian courts (Bala et al, 2010). The proposed research will make further contributions to the understanding and impact of PBRD instructions, and should contribute to the further development of standardized procedures and practices that can be adopted by the justice system. This program of research will significantly increase the ecological validity of our previous research. While the research is applied, it will further the theoretical understanding of the psychological factors that contribute to credibility assessments and understanding of forensic instructions.

## C. METHODOLOGY

The present research program consists of three inter-related projects.

**Project 1: Proof Beyond a Reasonable Doubt in the Justice System:** *Study 1.1* will analyze Canadian and international jurisprudence and scholarship with respect to PBRD instructions and will integrate psychological perspectives with legal analysis. This study will explore alternative articulations of the standard that will be used in later projects as well as resulting in publications.

**Study 1.2** will survey Canadian justice system professionals to assess their attitudes, application and understanding of PBRD instructions and related psychological issues. The study will build on our previous success in surveying Canadian judges and lawyers about child witness issues (Bala, Evans, & Bala, 2010; Bala et al, 2005). We will ask judges, prosecutors and defense counsel (N=180) for each group) questions related to PBRD instructions including common practices, subjective probability assessments, influence on assessments on cases, effectiveness of various instructions, and perceived jury understanding of such instructions. The survey will be conducted in two stages: a pilot survey in person, followed by a revised web-based questionnaire (via a password protected secure website) for targeted participants (with phone follow-up). We will use our existing database and presentations at professional education programs to recruit participants. The results of the survey will provide a context for generating hypotheses for our laboratory work and allow for appropriate interpretation of findings.

**Study 1.3** will examine the impact of PBRD instructions on the understanding of appropriate reasoning by jury-eligible laypersons (N=200). We will present a series of scenarios in which hypothetical jurors explain why they decided to, or not to, vote guilty based on their understanding of PBRD (e.g., “eyewitness can (can’t) be relied on to be accurate”);. Participants, given either (a) no definition or (b) current PBRD Canadian instructions, will be asked about whether the juror’s reasoning is an appropriate or inappropriate application of the PBRD criterion. We will compare participants’ ratings across scenarios and in the two instruction conditions.

**Project 2: Witness Assessments:** In court cases, “ground truth” is not known, but in experimental manipulations we can test PBRD instructions in the context of cases with known “right” and “wrong” answers. We will assess the effects of different PBRD instructions on the decisions of professionals and mock jurors in cases with adult and child witnesses in two of the most common types of issues in the criminal courts: (a) an eyewitness identification by a bystander, and (b) a victim’s report of an assault. Project 2 consists of 3 decision-making studies to examine accuracy of laypersons and justice system professionals in reaching verdicts when given different PBRD instructions. For the eyewitness identification stimuli for this project, children and adults will witness a simulated theft. These witnesses will then be presented with a photo line-up of suspects and asked to identify the perpetrator. Videos of both accurate eyewitness reports (i.e., the witness identifies the perpetrator) and inaccurate eyewitness reports (i.e., the witness mistakenly identifies someone else) will be used as stimuli. For the assault cases, adults and children will testify about true and fabricated reports of assault. To avoid ethical and methodological issues of creating or inducing criminal situations, we will study reports of peer-to-peer criminal acts (e.g., assault, robbery, extortion). This method has been used by us and others for adult and child subjects without harmful effects (Peterson & Parsons, 2005; Goodman, et al, 1994; Talwar et al, 2006).

**Study 2.1 Laypersons’ assessments of eyewitness reports of a theft:** Jury eligible participants (N=720) will be randomly assigned to either a full PBRD instruction condition or the no instruction condition. The effectiveness of different articulations of the PBRD standard (or no articulation of it) will be assessed by considering relative rates of guilty verdicts for accurate identifications, and not guilty verdicts for inaccurate identifications. Videos of “testimony” by adult and child witnesses will be viewed by adult participants to study the accuracy of adults’ assessment of witnesses’ accuracy. Strength of the evidence will be manipulated by presenting multiple witnesses either agreeing or disagreeing that the accused was the thief (based on pilot work to ensure a reasonable baseline “conviction rate” to test the hypotheses). A 2 (adult vs. child witness) X 2 (accurate vs. inaccurate IDs) X 2 (weak vs. strong evidence) X 3 (Control vs. PBRD instructions before evidence vs. PBRD instructions after evidence) design will be used. Adult laypersons will determine whether they believe the identified person was the criminal “beyond a reasonable doubt” and assess the likelihood (0% to 100%) that the identified person committed the crime. Demographic characteristics (age, sex, etc.) and beliefs of evaluators will be collected to examine their impact on the evaluators’ decision accuracy. The comparison of rated likelihood of the event when people did versus did not believe beyond a reasonable

doubt is expected to vary with presence vs. absence of reasonable doubt instructions to estimate the subjective probability of guilt associated with reasonable doubt (e.g. Dane, 1985).

**Study 2.2 Laypersons' assessments of assault reports.** Adult laypersons (N=360) will view videos of true or fabricated, children's vs. adults' reports of assault, and asked to decide with or without specific PBRD instructions whether they believe the events occurred "beyond a reasonable doubt" and to state the likelihood (0% to 100%) that the alleged events actually occurred. A 2 (adult vs. child witness) X 2 (true vs. false report) X 3 (Control vs. PBRD instructions before evidence vs. PBRD instructions after evidence) design will be used. Demographic characteristics (age, sex, etc.) and beliefs of evaluators will be collected. It is hypothesized that laypersons will be more accurate when provided with PBRD instructions and that they will be more likely to believe child witnesses (based on perceived honesty).

**Study 2.3 Justice system professionals' assessments of witness reports.** Similar to Studies 2.1 and 2.2, videos of testimony of child and adult witnesses will be viewed by judges and lawyers. Our team has experience and credibility in recruiting from this difficult to access subject pool. These professionals will be asked to determine whether the suggested events occurred or identifications were correct "beyond a reasonable doubt" and to state the likelihood (0% to 100%) that the witnesses' allegations are correct. They will also be given a short questionnaire on their beliefs about PBRD, the effects on decisions to prosecute/defend a case, and other questions generated from Project 1. The PBRD instruction manipulation will not be done here as the participants are assumed to have knowledge of the instructions. The manipulations of report content will be determined from the results of Studies 2.1 and 2.2, and only variables producing reliable effects in those studies will be manipulated in Study 2.3.

**Project 3: Deconstructing Beyond Reasonable Doubt Instructions** will test the judicial assumption (*R v. Lifchus*, 2007) that the entire PBRD instruction must be given to properly convey the meaning of PBRD. The 3 primary paragraphs of the Canadian PBRD instruction of *Lifchus* will be used in their entirety or omitting one of the paragraphs to determine the degree to which various subsections alter responses to the tasks described above. Project 3 will consist of 3 studies with laypersons.

**Study 3.1: Influence of portions of PBRD instructions on PBRD reasoning:** We will use the procedure identical to Study 1.3 except that adult laypersons (N=200) will be randomly assigned to one of the five conditions in which they are given the full Canadian PBRD instruction or with one of the three paragraphs omitted or no instruction at all. We will examine whether any part of the PBRD instruction plays an especially significant role in participants' judgments about various scenarios used in Study 1.3.

**Study 3.2 of portions of PBRD instructions on laypersons' assessments of eyewitness identification:** Similar to 2.1, adult laypersons (N=1200) will view the videos of eyewitness reports by children v adults that include accurate and inaccurate identifications about the identity of a thief and receive PBRD instructions. A 2 (adult v child witness) X 2 (weak v strong evidence) X 2 (accurate v inaccurate identification) X 5 (no instruction control, omission of the first paragraph, omission of the 2<sup>nd</sup> paragraph, omission of the 3<sup>rd</sup> paragraph, or complete PBRD instructions) design will be used. Measured variables will be identical to Study 2.1.

**Study 3.3 Influence of portions of PBRD instructions on laypersons' assessments of assault reports:** Adult laypersons (N=600) will view videos of reports of an assault and receive a portion of PBRD instructions, or no instruction. A 2 (adult v child witness) X 2 (true v false) X 5 (variation in PBRD instructions as above) design will be used. Measured variables will be identical to Study 2.2.

## **D. COMMUNICATION OF RESULTS**

Research and policy oriented papers will be submitted to Canadian and international journals in law and psychology. Our findings will be presented at national and international scholarly meetings, and at professional conferences, including judicial education programs sponsored by the National Judicial Institute; Bala, Dufaimont and Lindsay are regularly invited to present at NJI programs. All of the members of the research team have presented at national and international continuing education programs for lawyers, judges

and psychologists, often co-presenting on child witness issues, and will do so with these results. There will also be presentations for government policy-makers. Some of the conclusions will be of considerable public interest and we will issue news releases through media relations departments at our universities.