To be presented at, and to appear in the proceedings of, the IS&T/SPIE 16th International Symposium Electronic Imaging: Science and Technology, session on Optical Security and Counterfeit Deterrence Techniques V. Volume 5310 (January, 2004 San Jose, Ca)

Perception and Detection of Counterfeit Currency in Canada: Note Quality, Training and Security Features

Raymond M. Klein, Simon Gadbois, John J. Christie
Dalhousie University, Halifax, NS B3H 4J1, Canada

ABSTRACT

Commissioned by the Bank of Canada to help improve the detection of counterfeit currency, we designed a series of tests of performance to explore the contributions of note quality, sensory modality, training, security features and demographic variables to the accuracy of counterfeit detection with three different note types. In each test, participants (general public, and cash handlers, divided amongst commercial cash handlers and bank tellers) were presented with notes, one at a time, for up to seven seconds, and were asked to judge whether each note was genuine or counterfeit. With whole note inspection, overall accuracy was about 80%. When the security features were tested individually, the Optical Security Device (OSD) was the best feature, the hidden number was the worst, and the portrait, maple leaves, fluorescence, and microprinting were intermediate. Accuracy was higher with notes that could be seen but not touched than vice versa. Cash handlers were 74% correct with touch alone and adding touch to vision significantly improved counterfeit detection. This paper will demonstrate how performance differences between the different note types can be explained in terms of the efficacy of the individual security features incorporated into the notes.

Keywords: Banknotes, counterfeit, security features, perception, training, note quality

1. INTRODUCTION

Counterfeiting is a serious challenge to our confidence in the money we exchange on a daily basis. The Bank of Canada (BoC), like other central banks, has a mandate to discourage counterfeiting and to minimize its impact. Until recently the prevention of counterfeiting has focussed primarily on designing notes that make the counterfeiter's job as difficult as possible, a task challenged by the proliferation of inexpensive technology that can generate facsimiles of money at the touch of a button. Security features that are particularly difficult to duplicate have been developed to respond to this threat. Because there is a cost associated with each such security feature, to optimize resource allocation central banks seek to understand the effectiveness of these features . Whereas one aspect of "effectiveness" is technological and focussed on the counterfeiter ("How easy will it be for the professional or the opportunistic counterfeiter to duplicate this or that feature?"), the aspect explored here concerns the users of money With an ultimate interest in maximizing people's performance in detecting counterfeit currency, our research project with the BoC focussed on security features, note quality and training.

The BoC has funded many studies of attitudes and opinions collected from focus groups. But people's introspections (as collected through such self-reports) provide incomplete and sometimes inaccurate information about what they are attending and about what factors affect their performance (e.g., Berry, 1996). Studies of actual user performance at counterfeit detection are therefore needed to provide more objective answers to such questions. The need for such data on perception and performance has been recognized before (Croney, 1970; 1974), and reflecting this need, the United States Treasury recently commissioned a research project on counterfeit detection performance (Hillstrom & Bernstein, 2002). That study was thorough and competently executed. Nevertheless, before applying a similar approach to Canadian currency we made one important improvement. Because the US study used a "forced-choice" methodology the task was more akin to asking "which of these two notes is more genuine" than to asking “is this particular note genuine or counterfeit”. The forced choice method is common in perception research and is well-suited for discovering the limits of perception: How small a difference can be detected? However, for the purpose of generating performance data with relevance to the kind of real-world monetary transactions that concern the BoC, knowing the best case performance may be misleading and is unrepresentative of counterfeit detection in general. Therefore, in our study, notes were presented to individuals one at a time. This way, as in everyday transactions, each note is treated as an individual one that may, or may not, be counterfeit.

2. METHODOLOGY

2.1. Participants

A total of 158 participants were recruited: Seventy-nine were members of the general public (GP) who did not manipulate money regularly as part of their job, and seventy nine were cash handlers (CH) who did manipulate money regularly as part of their job. CH were divided into roughly equal groups of individuals who either worked in a financial institution, such as chartered bank or credit union (Bank Tellers, BT; n=39) or in a retail environment (Commercial Cash Handlers, CCH; *n*=40). Participants in each main group (GP or CH) were randomly assigned to three training conditions: No Training, Leaflet Training or Video Training (see below and Figure 1). 17 of the participants were Francophone. Of these, 16 were GP, who were as evenly divided amongst training conditions as possible.

2.2. Notes

A total of 848 notes were used, of which 492 were genuine and 356 were counterfeit. These notes were drawn from the Birds of Canada series ($10s and $20s, B10, B20), first introduced in 1986, and the Canadian Journey $10 (J10), the first of a new series (see http://www.bankofcanada.ca/en/banknotes/). The notes were selected from three quality levels that we will refer to as low, medium and high, and in each step, counterfeit and genuine notes were matched for quality.

2.3. Procedures

Participants were welcomed in the testing room and, after introductions, were immediately asked the following question: “What is the first thing you notice on a note during a cash transaction?” Informed consent was next obtained from all the participants. Electronic versions of two vision tests were administered using iBook computers to assess colour blindness (Ishihara) and visual acuity (Snellen). When tested with any corrective lenses they may have brought with them, 23 individuals had one vision problem or another. These participants were not excluded as we considered that their presence enhanced the representativeness of our sample. Our findings were not significantly affected by this decision. At this point, a sequence of steps began that are described below (see Figure 1). In an effort to minimize the contribution of varying expectations participants were informed of the actual proportion of counterfeits used in each Step.

Step 1: Whole note test

Participants were asked to examine a series of notes for a period of no more than 7 seconds each. A tone generated by the computer signalled the beginning and, 7 seconds later, a second tone signalled the end of the inspection period when participants were required to return the note (if they had not already done so) and judge if it was genuine or counterfeit. 168 notes were divided into two matched sets (A and B) with 1/2 the participants in each subgroup getting set A and 1/2 set B in this step. Thus, each participant responded to 84 notes, 1/3 of which were counterfeit and 2/3 genuine. The 84 notes in each set were presented in

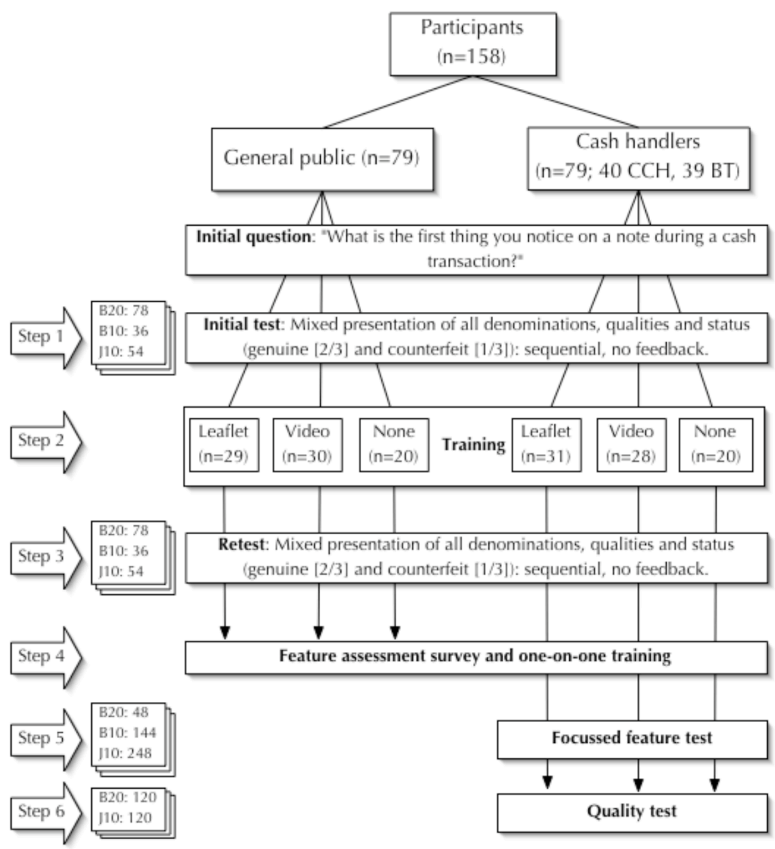

Figure 1. Flow chart illustrating the sequence of steps that participants in different groups experienced in the experiment. The total number of notes that was available for each stop is shown in the boxes to the right of the Step-arrows

one of 4 pseudo-random orders to minimize practice and fatigue effects across notes.

Step 2: Training
Participants were exposed to training material prepared by the BoC's Bank Note Communication and Compliance Team (video or leaflet; 59 GP and 59 CH) or to a control activity (rest or read; 20 GP and 20 CH) for 8 minutes, (the length of the video). For the participants exposed to training material, roughly half were presented with a digitized version of a video and the remainder with leaflets. The language of the training material (video or leaflet) was French for the francophone participants; otherwise it was English.

Step 3: Whole note retest
Step 3 was a repetition of Step 1, with the alternate set of notes. The sharing of notes between steps 1 and 3 ensures that comparisons of Steps 1 and 3 to assess the effect of training are based on precisely the same notes and provides a large and useful data-base:158 decisions on 168 carefully selected notes of different quality levels, denominations and series.

Step 4: Feature assessment survey
This step began with a survey to determine which features of the notes the participants thought they had been using to make their judgements in Steps 1 and 3, and ended with a one-on-one training session to ensure that participants understood how to use each of the 7 key features of the notes in our study (OSD, Micro Printing, Portrait, Fluorescence, Iridescent Maple Leaves, Hidden Number and Raised Ink). A UV lamp was provided to illustrate Fluorescence and a magnifying glass was provided for the Micro Printing and Portrait. The interested reader is referred to the BoC's web site for an explanation of these features:
http://www.bankofcanada.ca/en/banknotes/counterfeit/security/index.html
Only the CH proceeded on to Steps 5 and 6.

Step 5: Focussed feature and modality tests
Nine subtests designed to meet two related objectives were interleaved in Step 5. As in Steps 1 & 3 notes in each subtest were presented for up to 7 seconds with genuine/counterfeit judgements required for each one. One objective was to ask, for each of the main "visual" security features (OSD, Micro Printing, Portrait, Fluorescence, Iridescent Maple Leaves and Hidden Number): "How well does this feature perform when it is the only one available?" In each of six "focussed feature" subtests only the feature of interest could be perceived; the rest of the note was masked by black paper and inserted in a plastic sleeve to eliminate any tactile cues. As needed, one of two tools (that were demonstrated in Step 4) was provided. For the Fluorescence subtest, the UV lamp was available and the lights in the experimental room were turned off; for the Micro Printing and the Portrait subtests the participants could use the magnifying glass. The other objective was to compare inspection of whole notes using vision alone (V), touch alone (T) and vision and touch combined (VT). In the V condition, the participants evaluated the notes as they did in Steps 1 & 3 (i.e., in full), except that the notes were in plastic sleeves so the participants could not feel them. In the T condition, participants were allowed to touch the notes as in Steps 1 and 3, but a screen was used to prevent vision of the notes. Finally, in the VT condition, participants could touch and see the notes exactly as in Steps 1 and 3. The order of the 9 subtests was counterbalanced across participants to ensure that each feature was tested, equally often across participants, at the same levels of practice. For the modality tests, each specific note was used equally often in each of the three presentation conditions. For all tests in this step, the proportion of counterfeit to genuine notes was 1:1.

Step 6: Quality test
Step 6 assessed performance using high and low quality notes that were embedded in primarily high and low quality contexts. Participants were asked to sort, into genuine and counterfeit piles, each of two stacks of notes. Each stack was composed of a random mix of genuine (2/3rds) and counterfeit (1/3rd) exemplars of the J10s and B20s. In the "high quality" stack, 80% of the notes were of high quality while the remaining 20% were of low quality. In the "low quality" stack the proportions were reversed. Participants were asked to work as fast and as accurately as possible. Speed and accuracy were the dependent variables. Assignment of individual notes to the different stacks was counterbalanced across participants such that regardless of its quality, each note was present in the high and low quality stack.

3. RESULTS & DISCUSSION

Percent correct, for the counterfeit and genuine notes and for both types combined, is the primary measure of performance. With this measure, higher scores will consistently correspond to a better performance (correct identification of a genuine note when it is in fact a genuine, correct identification of counterfeit when it is fact a counterfeit). Unless otherwise noted all effects and differences that are explicitly described are significant using a conventional and appropriate statistical procedure, e.g. t-test, ANOVA, etc., with alpha set at .05.

3.1 What people say they notice first

In response to the initial question: “What is the first thing you notice on a note during a cash transaction?” over 70% of the GP indicated a direct (“denomination” or “number”) or indirect (“colour”) effort to determine the note’s denomination. A remarkably small number of the GP (12.5%) specifically mentioned a security feature. Perhaps because professional CH are alert to the possibility of counterfeiting, a smaller percentage of these participants answered with a “denomination-seeking” response (40% for CCH and 28% for BT). Specific security features or touch was mentioned by over 50% of the participants in each of the two CH groups with the OSD mentioned most frequently (about 20%).

3.2 Step 1: Whole note test before training

Performance in this step provides untrained performance levels of GP and CH with different note types (B10s and B20s and J10s) varying in quality. In their overall performance, BT (87% correct) were better than CCH (81%), who, in turn were better than the GP (74%). It should not be surprising that individuals who handle money as part of their job would perform more accurately than individuals who do not.

Performance varied with type of note and with note quality for which there were three levels (low=1 , medium=3 and high=5), and the effect of quality was different for the different note types. Overall performance with the B10s and B20s (80.5% & 81.2%, respectively) was more accurate than with the J10s (75.7%). Quality had somewhat different effects on the classification of counterfeit and genuine notes, but interpretation is complicated by the fact that the pattern also depended on note type. When a B10 note was high in quality, there was a strong tendency to classify it as counterfeit. This bias produces what appears to be a high counterfeit recognition of the notes from this series, (97%) combined with a low performance with the genuine notes (71%). With the B20, there was a weaker tendency for high quality genuine notes to be classified incorrectly as counterfeit (87% compared with 91% for low and medium quality combined). This was not a bias because, unlike the pattern with the B10s, performance on the high quality counterfeit B20s was not exceptionally high (73%). In contrast, with the genuine J10s, it was the low quality notes that were most likely to be classified incorrectly as counterfeit (78% vs 86% for medium and high quality). Whereas the patterns of performance with the B20s and J10s did not vary significantly with group, the pattern with the B10s did. The GP mistakenly classify almost 40% of the highest quality (level 5) genuine B10s as counterfeit. In contrast, their classification of the medium and low quality genuine notes was much more accurate. Comparison with performance on the counterfeit notes reveals that these differences are not so much about the ability of GP to distinguish counterfeit from genuine as a function of quality but rather the pattern indicates that GP tend to believe that high quality B10s are counterfeit and that low quality notes from this series are genuine. Whereas this bias to think that crisp new B10s are counterfeit (and conversely used and soiled B10s are genuine) is present, though to a lesser degree in the CCH, it is more or less absent in the BT. Perhaps this can be explained by the fact that fewer and fewer B10s (especially in crisp new condition) are in circulation since the introduction of the new J10s in 2001 (two years before this study was conducted). Because the typical lifetime of a $10 note is one year, it is safe to assume that the B10s that were circulating at the time of the study were mostly low in quality. Therefore, in their recent experiences, nearly all B10s that the GP and CCH encounter would have been low quality notes. We suggest that because the low quality notes in this study were similar to what was in circulation, they tended to be classified as real; the high quality notes were dissimilar and tended to be classified as counterfeit. It is noteworthy that the BT were not subject to this bias. This could be because their training or job makes them more reflective, or perhaps because their increased experience with high quality notes generalizes to the B10s.

3.2 Step 3 minus Step1: Effects of Practice and Training

To explore the effect of training two baselines must be considered. First, each participant’s performance in Step 1 is subtracted from that in Step 3 to determine the amount of improvement between steps (see Table 1). Second, the subgroups that received no training during Step 2 provide an index of how much improvement one might expect to see merely from continued practice on the task. The improvement scores for the group that received no training reveal a significant improvement (approximately 4%) in Step 3 that can be attributed to practice on the task and exposure to counterfeits. In order to assess which training experience (leaflet or video) produced more improvement than no

	Training		
Group	**None**	**Leaflet**	**Video**
BT	4.3	0.7	3.4
CCH	4.8	5.1	9.9
GP	3.1	4.9	5.5

Table 1. Improvement (accuracy: Step 3 minus Step 1) for each combination of group and training collapsed across counterfeit and genuine notes (in %).

training at all, the data were analysed separately for each group, because the groups appear to differ in their response to training. Strikingly, the BT did not benefit at all from either training tool, showing as much if not greater improvement with no training. This effect may be related to the relatively high skill levels amongst BT who have well-developed habits which mediate their counterfeit detection performance. Encouraging them to think about what they were doing (Herrigel, 1953) and look for particular features (Reber, 1976) may have interfered with these habits. With the CCH, the video produced the greatest performance improvement with both genuine and counterfeit notes. Although the benefit from the video (compared with no training) was not statistically significant, it must be kept in mind that due to the relatively small number of CCH in each of the three subgroups of CCH (8, 12 and 20, in the none, video and leaflet conditions, respectively, of Step 2) there may be a true benefit from the video that we do not have sufficient power to detect. For the GP exposure to the video produced non-significant improvements, with both counterfeit and genuine notes, compared to no training. Exposure to the leaflet produced significantly better counterfeit identification than did no training. But, because exposure to the leaflet also produced a small (nonsignificant) reduction (compared to no training) in the identification of genuine notes, one component of the effect of exposure to the leaflet is to make this group more suspicious.

To further explore the effects of training and practice, we looked at the different types of notes. One important finding related to practice was the considerable improvement that CH showed regardless of training in the identification of counterfeit B20. From Step 1 to Step 3, accuracy without training improved by 9.3% for the BT and 19.2% for the CCH on counterfeit B20s. These practice effects were not much affected by training as the improvements for the groups with training (video and leaflet combined) were 7% and 21% for the BT and CCH, respectively. We believe that once experienced cash handlers begin to see poorly reproduced OSDs they become very good at using this feature. For the GP, the benefits of training were most marked for the B20s. With counterfeit B20s, the leaflet produced a 15% improvement and the video produced an 11% improvement while the improvement with no training was only 5%. With genuine B20s, the video produced more improvement (7%), than the leaflet and no training (3% and 1%, respectively).

3.3 Steps 1 and 3 Combined

As noted combining the results from Steps 1 and 3 provides an excellent database for exploring how different note properties affect performance (see Table 2 for breakdowns into finer categories than presented below).

We first explored accuracy with note type (B10s, B20s and J10s) as a factor. The results are shown in Figure 2. With the counterfeit notes the J10s produced significantly worse performance than either of the Birds of Canada notes which did not differ. Analysis of the genuine notes revealed significantly more accurate performance with the B20s than either the B10s or the J10s which did not differ. When considering overall performance, participants were most accurate with the B20s and least accurate with the J10s. One might think that the poor performance with the J10s is attributable to their relatively recent introduction, and therefore the lower amount of experience participants might have with these notes. This seems unlikely, however, in light of the fact that performance on this note is the poorest for all three groups, including the BT, who probably have the most experience with them. Instead, we think that the J10s perform poorly here because they have been more successfully duplicated than the B10s or $20s. There are several reasons why this might be the case. First, if it is the case that the OSD is difficult to counterfeit (see Step 5), then it would be logical for

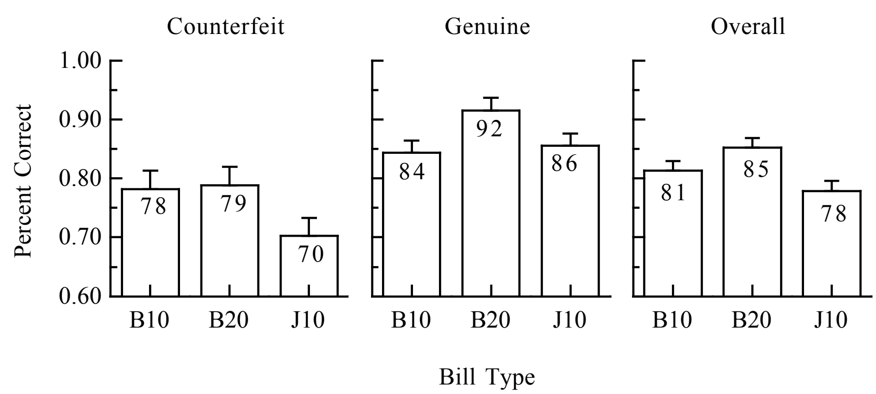

Figure 2: Accuracy in Steps 1 & 3 by note type, plotted separately for counterfeit and genuine notes, and for counterfeit and genuine combined. Comparisons within a panel, to determine if performance is affected significantly by the factor plotted on the X axis, are facilitated by the use of error bars (derived from the least-significance difference test) that will be referred to as “comparison bars” to avoid confusion with other kinds of bars (standard error bars; confidence intervals; etc.) and which will be used in figures 3, 4 and 6. When two conditions differ by more than the length of the comparison bar for that panel, the difference between the two conditions is statistically significant.

		Counterfeit				Genuine				Overall			
Group	**Quality**	**J10**	**B10**	**B20**	**All**	**J10**	**B10**	**B20**	**All**	**J10**	**B10**	**B20**	**All**
CCH	1	78	77	84	**80**	80	91	94	**88**	79	84	89	**84**
	3	72	69	83	**75**	90	89	95	**91**	81	79	89	**83**
	5	72	97	83	**84**	93	82	90	**88**	82	90	87	**86**
	All	**74**	**81**	**84**	**80**	**88**	**87**	**93**	**89**	**81**	**84**	**88**	**84**
BT	1	82	85	88	**85**	84	88	94	**89**	83	87	91	**87**
	3	77	86	89	**84**	90	95	96	**94**	84	90	92	**89**
	5	77	96	89	**88**	94	90	95	**93**	85	93	92	**90**
	All	**79**	**89**	**89**	**85**	**89**	**91**	**95**	**92**	**84**	**90**	**92**	**89**
CH	1	80	81	86	**82**	82	90	94	**89**	81	85	90	**85**
	3	75	77	86	**79**	90	92	95	**93**	82	85	91	**86**
	5	75	97	86	**86**	93	86	92	**90**	84	91	89	**88**
	All	**77**	**85**	**86**	**83**	**88**	**89**	**94**	**91**	**83**	**87**	**90**	**87**
GP	1	65	63	71	**66**	76	87	90	**84**	70	75	80	**75**
	3	61	57	70	**63**	85	85	91	**87**	73	71	80	**75**
	5	65	95	75	**78**	87	67	87	**80**	76	81	81	**79**
	All	**64**	**72**	**72**	**69**	**83**	**80**	**89**	**84**	**73**	**76**	**80**	**76**
All	1	73	72	78	**74**	79	88	92	**86**	76	80	85	**80**
	3	68	67	78	**71**	88	88	93	**90**	78	78	85	**80**
	5	70	96	81	**82**	90	76	89	**85**	80	86	85	**84**
Total:	**All**	**70**	**78**	**79**	**76**	**86**	**84**	**92**	**87**	**78**	**81**	**85**	**81**

Table 2. Accuracy (%) in Steps 1 and 3 combined as a function of note type, group and quality level.

counterfeiters to invest their resources in notes, like the J10s, which do not have this troublesome feature. Second, there was less Raised Ink (intaglio) on the J10s than on the B10s. Intaglio is relatively difficult to duplicate and was not present on the counterfeits used in our study. Because the difference in tactility between the genuine and counterfeit J10s was smaller than that for the Birds notes, the B10s and B20s might have had an advantage to the extent that participants pick up useful information from handling the note. For converging evidence, see 3.6 Step 5 Modality tests.

We next considered performance as a function of note quality. With counterfeits, performance with high quality notes was significantly better than with low or medium quality notes, which did not differ. Although the differences were small, with genuine notes accuracy was significantly higher with medium than with either low or high quality notes, which did not differ. It might be tempting to conclude from the relatively high rate of counterfeit classification with the high quality notes (82%) that it is easier to identify counterfeit notes when they are of high quality. The somewhat lower level of performance with the high quality genuine notes (high = 85% compared with medium = 90%) suggests a different possibility that cannot be ruled out: There is a tendency to be somewhat suspicious of new notes. In the Step 1 analysis there was a remarkably strong version of such a bias when the GP were judging B10s. Indeed, the relatively good performance with high quality counterfeits was almost completely due to the B10s. Nevertheless, in the overall (counterfeit and genuine combined) data, performance is best with the high quality notes (84% vs 80%).

3.4 Step 4: Survey

When participants were asked, of the three notes, “which did you find easiest to authenticate”, the J10s and B20s were preferred over the B10s. When those who preferred the B20s were asked to choose, among the two series of $10s, about 60% picked the J10s. Thus, when only the $10s are considered 70% thought the J10s were easier to authenticate than the B10s. As already noted, this was not actually true of performance which revealed that our participants performed more accurately with the B10s than the J10s. There were no differences between the GP and the CH in their impression that the B10s were the hardest to authenticate. After participants answered: “Which elements of the note did you use to make your decision as to whether it was genuine or counterfeit?” we asked them: “Out of the elements you just chose, please rank these in order of most important to least important in determining whether a note was genuine or counterfeit.” We were interested in the frequency with which different features were selected at this stage of the experiment as the “most important” for counterfeit detection. Among those who had not received training in Step 2 over 30% of the participants mentioned the OSD and over 30% mentioned Touch. Training increased reports of the Iridiscent Maple Leaves and Raised Ink (substituting for some of the simple Touch responses).

3.5 Step 5: Focussed feature tests

The data from the focussed feature tests are shown in Figure 3, collapsed across the BT and CCH groups because the relative ordering of performance (from best to worst) of the different features was the same for the two groups. Particularly noteworthy is the relatively high accuracy in identifying counterfeits using the OSD and relatively low accuracy on both counterfeit and genuine notes when processing was restricted to the Hidden Number (the remaining

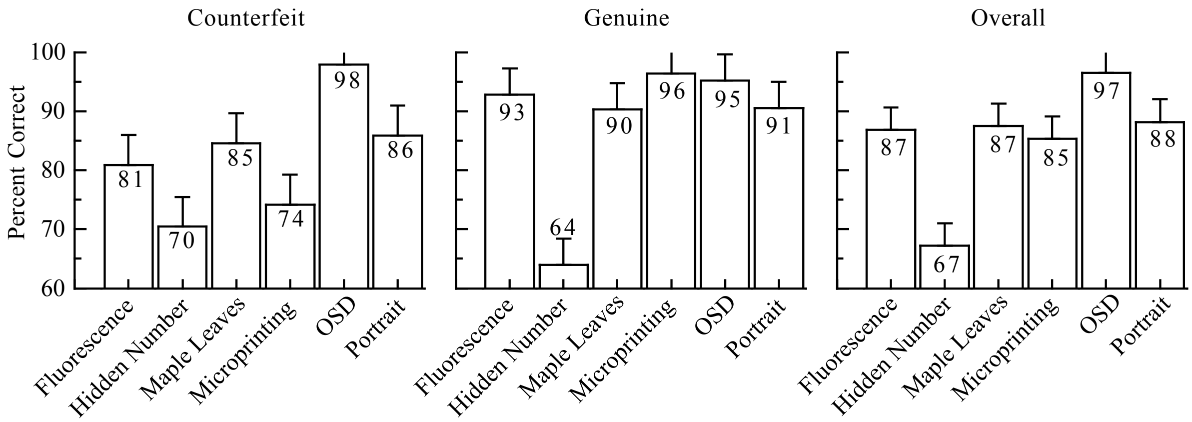

Figure 3. Accuracy as a function of the feature that participants were constrained to use in Step 5.

features performing at a similar level between these extremes). We believe that the OSD performs the best overall because it is a perceptually salient feature that is difficult to duplicate well.

Although reliance on Micro Printing alone produced relatively poor performance with counterfeit notes, it yielded the highest level of performance with the genuine notes. One possible explanation for this pattern is that this is a feature that counterfeiters are duplicating reasonably well. With perception restricted to relatively good Micro Printing (on both genuine and, according to this suggestion, counterfeit notes), participants would correctly classify genuine notes as such and would tend to misclassify counterfeits as genuine. Alternatively, perhaps when restricted to this feature, participants were merely biased to make a "genuine" classification. Support for the former alternative would require measurement of the quality of counterfeit reproductions of Micro Printing, combined with psychophysical assessment of observers' capacity to perceive any differences found.

The new J10 note shares with its older counterpart, the B10 note, three features: Fluorescence, Micro Printing and Portrait. Comparison of these features across the two series (see Figure 4) reveals several striking findings. In the

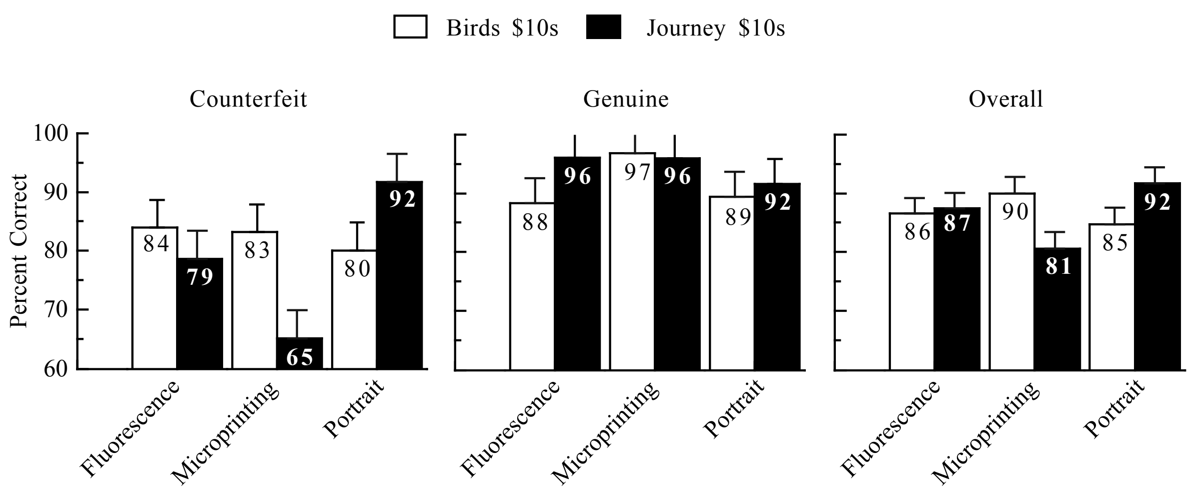

Figure 4. Accuracy with counterfeit and genuine notes for the three security features shared by the old (Birds of Canada) and new (Canadian Journey) series of $10s.

overall analysis there was no difference between the two series with processing confined to Fluorescence. Yet looking at the counterfeit and genuine notes separately reveals that there was a tendency for participants to say “genuine” with the J10s. With Micro Printing performance was much better with the counterfeit B10s than with the counterfeit J10s, with twice as many erroneous genuine judgements for the latter notes (35% versus 17% errors). With the Portrait this effect is reversed, with the J10s showing a large advantage over the B10s (20% versus 8% errors). These effects are not present with the genuine notes. Perhaps this means that the new Portrait is hard to counterfeit while the Micro Printing is either easy to duplicate or hard to process. Whether the J10s’ Portrait is better because of its increased size, colour or the additional fine lines, or some combination of these changes, we can’t say. We believe that the graded Micro Printing on the J10s (see Figure 5) may not be an effective security feature. People may tend to look for the gradation and may not focus on how well the details are preserved at the smallest scale. Regardless of viewing conditions and the relative success in reproducing this feature on a counterfeit, some gradation will be noticeable. Therefore, perceiving a relative gradation would not be a useful cue for authenticating a note. This may not be true when magnification is used. Although we provided a magnifying glass that might have been used in the Micro Printing subtest, data on the frequency of its usage during this subtest is not available.

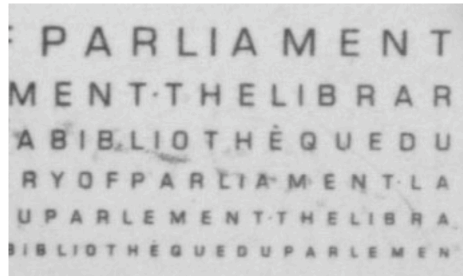

Figure 5. Graded microprinting on the J10.

The authentication of some individual counterfeit notes was as low as 15% correct when processing was confined to a single feature. This highlights the importance of the use of multiple security features in currency design. Otherwise, counterfeiters may be able to succeed by focussing their resources on just one or two features. Conversely, it is equally important for those handling money (GP and CH) to be well-educated about these features and to allocate their attention among them. To illustrate this last point, suppose that multiple features are incorporated into a note series, and, in addition, suppose that when authenticating these notes people focus their attention on the same, single feature. In effect, this behaviour would create a niche in which the counterfeiter, by focussing resources on reproducing the attended feature, can be successful even though, as a whole, the counterfeit note may be a poor reproduction. Thus the integration of multiple features in the design of notes is necessary, but not sufficient, to minimize successful counterfeiting. It must be accompanied by awareness of and attention to these multiple features by those who handle the currency, particularly by those who might be recipients of counterfeit notes.

3.6 Step 5 Modality tests: Vision, Touch and Vision+Touch

Modality results are shown in Figure 6. Overall, there was much better performance with notes that could be seen but not touched (87%) than vice versa (74%). With the counterfeit notes adding touch to vision improved performance, which was not the case with the genuine notes. It seems likely that the visual properties that can be used to distinguish genuine from counterfeit notes are more highly discriminable than the tactile properties that can make this distinction. Knowing the denomination and series is important for tactile inspection of the notes, and with touch alone it would be difficult to make this determination.

Our experimenters reported anecdotally that in the Vision plus Touch condition some participants explored, via touch, notes from the different series in a different manner. In the modality tests, because the different series were randomly intermixed, without vision participants would not easily know the note type. It seems likely, therefore, that they might have adopted a strategy that was appropriate for only one series; or they may have needed too much time to determine what series they were examining, leaving insufficient time to find the telling tactile information. In contrast, with vision plus touch, participants may have used touch more effectively since, while exploring specifically visual features, the participant would have quickly established the series and hence might have known where to touch the note for the Raised Ink or Tactile Feature. Analysis of the videotaped behavior

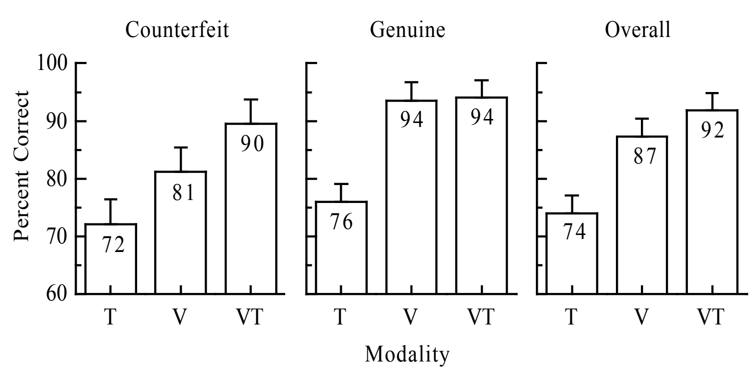

Figure 6. Accuracy as a function of the modality used to inspect the note (T=touch, V=vision, VT= vision and touch).

from this step (Gadbois, Klein & Christie, 2004) support this conjecture.

The counterfeit J10s were exceptionally difficult to correctly classify using touch alone, resulting in a particularly poor performance (55% correct, compared to 76% and 85% for the B20s and 10s, respectively). This corroborates our earlier suggestion that the relatively poor performance of the J10s in Steps 1 & 3 (70% correct with counterfeits) might be attributed to the relatively unsalient amount of intaglio on the genuine notes in this series. The low tactility of the J10s in circulation at the time of our study was recognized by the BoC, which issued new J10s with increased intaglio in January, 2003 in an effort to improve these notes.

3.7 Step 5: Features and modalities compared

In typical cash transactions, notes are perceived using vision and touch combined. It therefore seems appropriate to use performance in this condition as a yardstick against which to assess the relative efficacy of the security features tested in the Focussed Feature tests. Yet CH may be encouraged to use a particular tool (such as a UV lamp) which focusses their processing on a particular feature, or their training may emphasize some features at the expense of others. Does such focussed processing help or hurt performance? The overall data (counterfeit and genuine combined) from the focussed feature and modality subtests have been combined in Table 3 to facilitate comparison of performance using individual features or the whole note. Because the data shown here is based on relatively small samples of selected notes the interesting differences that emerge must be regarded cautiously. Generally, and as might be expected, performance was better when the whole note was inspected with vision and touch than when processing was limited to a single feature. There are two noteworthy exceptions to this generalization. Performance using either the OSD or J10 Portrait alone was numerically superior to that with whole note inspection using vision and touch. This reinforces two conclusions drawn previously: The OSD is a standout feature for the purpose of counterfeit identification and, despite the overall poor performance of the J10s, its Portrait performs quite well and should be emphasized.

	Series		
Feature	**B20**	**B10**	**J10**
OSD	97		
Micro Printing		90	81
Fluorescence		86	87
Portrait		85	92
Iridescent Maple Leaves			87
Hidden Number			67
Touch only (T)	74	82	66
Vision only (V)	92	86	83
Vision + touch (VT)	94	94	88

Table 3. Overall accuracy for the different series of notes in all the subtests of Step 5.

3.8 Step 6: Quality test

In rapidly sorting stack of notes into genuine and counterfeit piles, the BT were significantly faster than the CCH and both groups were faster when the quality of the stack was mostly high than when it was mostly low. Low quality counterfeit notes were more accurately identified than the high quality counterfeits, an effect that was due entirely to the J10s (88% correct for low quality and 82% correct for high quality). The quality context had no effect on accuracy with counterfeit notes. With the genuine notes, performance was, again, more accurate for the B20s than for the J10s and more accurate for high than for low quality notes. Quality context had no effect on classification accuracy with high quality genuine notes, but low quality genuine notes were less accurately identified in the high quality context. In other words, low quality genuine notes were more accurately classified when the quality context was mostly low than mostly high. This effect of context was due entirely to the low quality J10s. Accuracy for the low quality J10s was 11% worse when most of the other notes were high in quality (76% correct) than when most of the other notes were low in quality (87% correct). Because performance is generally better with high, than with low quality notes, particularly genuine notes which make up the vast majority of notes we handle, it is certainly a beneficial strategy to remove the lower quality notes from circulation. The findings from this step suggest that this is particularly important with the J10s because as they decrease in quality they suffer greatly when judged in the context of mostly high quality notes.

3.9 Some additional findings

In Step 4 participants were asked to give a 1-10 rating of the degree to which they relied upon “the feel or touch” or the “look” of the notes when making their genuine/counterfeit classifications. To see whether these self-reported degrees of reliance upon touch and vision predicted how well participants performed when they could use only touch and vision, respectively, we calculated the correlation between the survey responses to these questions, and performance from the touch and vision tests (respectively), of the Modality subtest of Step 5. The degree to which CH said they relied on touch was not related to performance when only touch was used in the Modality substep of Step 5 ($r=-.055$), nor was it related to the amount of improvement in performance that CH showed when touch was added to vision (VT-

V, r=-.123). Similarly, the degree to which CH said they relied upon the look of the note was not related to performance in the vision test (r=.076), nor was it related to the amount of improvement in performance that they showed when vision was added to touch (VT-T, r=.114). To put these low correlations in perspective, with n=79 a correlation would have to be .23 or greater to be considered significant. Sometimes a true relationship is not significant because there is insufficient variation in one of the two measures. The failure of self-reported use of vision and touch to be significantly related to performance when processing was confined to vision and touch is not, however, due to insufficient variation: Touch ratings ranged from 0 to 10; Vision ratings ranged from 3 to 10; overall performance with touch ranged from .42 to .92 while overall performance with vision ranged from .53 to 1.0. The most likely interpretation is that self-reports are unreliable indicators of what our participants actually used when making genuine/counterfeit judgements.

We also examined the relationship between the answers to the question what aspect was “most...important in determining whether a note was genuine or counterfeit?”) and performance for all the pertinent individual features tested in Step 5 (OSD, Hidden Number, Iridescent Maple Leaves, Micro Printing, Portrait and Touch). It seems reasonable to expect that participants who ranked a specific feature as most important in Step 4 (after making their judgements in Steps 1 and 3) would perform particularly well on this feature test in Step 5, compared to the other participants (who preferred a different feature). Generally speaking, this is not what was found (Figure 7). Performance by those ranking a feature as “most important” (filled circles) was generally well within the range shown by most of the CH when processing was confined to that feature. There was a tendency for those ranking the Iridescent Maple Leaves as “most important” to perform slightly worse than the mean for the group as a whole (open circles) using this feature (82% vs 87%) and for those ranking Micro Printing and the Portrait as “most important” to perform slightly better than the mean for the group as a whole (96% vs 85% for Micro Printing, and 98% vs 88% for Portrait). It is tempting to consider whether these data suggest that training should emphasize Micro Printing and the Portrait. Indeed, performance by those preferring Micro Printing and the Portrait when confined to their preferred feature is as accurate as when performance is confined to the most accurate OSD. With only 5or 6 participants endorsing these features, however, the pattern shown in Figure 7 must be regarded as suggestive.

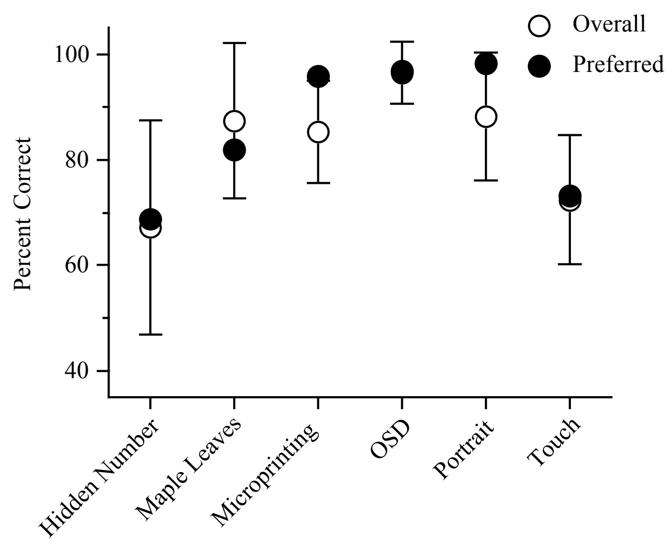

Figure 7. Overall (counterfeit and genuine combined) accuracy in Step 5 when processing was confined to a particular feature or touch shown for all CH participants (open circles=mean) along with those who, in Step 4, endorsed the plotted feature as the one that was “most important” in making their judgements (filled circles=mean). The bars in this figure represent +/- 1 standard deviation of the mean proportion correct for all CH. Thus, roughly two thirds of the CH’s scores, when processing was confined to each note feature shown, would fall within the range indicated by the bar. The number of participants endorsing the plotted features (and thus contributing to each filled circle, respectively) were: Hidden Number = 4, Iridescent Maple Leaves = 6, Micro Printing = 6, OSD = 26, Portrait = 5, any aspect of Touch = 21.

These results show that experimentally measured performance on counterfeit detection is more important than the perceived performance or reported attention to specific features. In other words, self-reported usage of features (as might be collected from surveys and focus groups) is not a satisfactory index of the actual discrimination and detection behaviour of the participants. Survey data should be interpreted with caution since it is based on self-reports which may not reflect actual performance, feature usage or feature efficacy.

6. SUMMARY, CONCLUSIONS & RECOMMENDATIONS

Overall performance authenticating these notes was about 80%, which compares favourably with US and British studies using similar methods. BT performed better than CCH who in turn were better than the GP. Performance with the B10s and 20s was better than with the J10s. Initially, performance on the B10s and B20s did not differ; after some practice on the task, and we suggest because participants learned to use the highly effective Optical Security Device (OSD), participants became most accurate with the B20s. Several findings support our proposal that the poor performance of the J10s is rooted in its low tactility (insufficient intaglio ink).

Overall accuracy was best with high quality notes. Furthermore, before receiving any training on counterfeit detection, the GP (and to a lesser degree the CCH) were biased to classify high quality notes from the discontinued B10 series as counterfeit, a tendency that continued in weaker form after experience with the task. It seems reasonable to predict that this would be a general finding with notes from a previously discontinued series so long as those remaining in circulation are mostly low in quality.

Performance of the GP on counterfeit note detection benefited from training, with the leaflet contributing more than the video. CCH showed little (video) or no (leaflet) benefit from training, while the performance of the BT, in part because of their already high level of performance and prior experience, showed a slight decrease in performance. It is important to keep in mind that the BoCs' educational materials were assessed after participants had made 84 genuine/counterfeit judgements. It is possible that more positive effects of these materials would have been observed were they assessed <u>before</u> exposure to a task like that in Step 1. Also, during Steps 1 and 3 participants received no feedback on the accuracy of their responses. Known principles of learning (Trowbridge & Cason, 1932) predict that the practice effects we obtained might be enhanced with such feedback. Because of the robust practice effect, we recommend the development of a training protocol that combines exposure to actual counterfeit notes with practice making genuine/counterfeit judgements in the presence of feedback.

When specific visual features were individually tested, the OSD produced the best overall performance and the Hidden Number produced the lowest performance, with the other features (Portrait, Iridescent Maple Leaves, Fluorescence and Micro Printing) performing at a similar level between these extremes. An interesting finding emerged when comparing the security features that are common to the two series of $10 notes: Micro Printing on the B10s outperformed that on the J10s, while the J10s' Portrait was superior. Whether the J10s' Portrait is better because of its increased size, colour or the additional fine lines, or some combination of these changes, we can't say. We believe that the graded Micro Printing on the J10s may not be an effective security feature.

When the sensory modalities were compared performance was much better with notes that could be seen but not touched than vice versa, although accuracy with touch alone was a respectable 74%. Adding touch to vision improved counterfeit detection but did not affect performance with genuine notes. We suggest that due to the relatively low level of intaglio on genuine J10s, counterfeit J10s were exceptionally difficult to detect using touch alone, resulting in a particularly poor performance (55% correct).

When the speed of classifying notes as genuine and counterfeit was measured using stacks of notes that were primarily high or low in quality, we found that BT were much faster than CCH, and both groups were faster with the high quality stacks. The quality context (mostly high or mostly low in quality) affected the identification accuracy of low quality genuine J10s (counterfeits, high quality genuine J10s, and B20s were unaffected by context). When most of the other notes were low in quality, 13% of the low quality genuine J10s were erroneously considered counterfeit whereas when embedded in a high quality context the error rate increased to 24%. Since the BoC is already trying to maintain a generally high quality context, it is particularly important to remove low quality J10s from circulation.

With two exceptions (OSD and Portrait), performance was worse when participants could only examine a specific security feature than when they could examine the whole note; in the focussed feature tests, the performance of the OSD on the B20s was outstanding and no counterfeit note generated more than 5.3% errors. The second best performer was the Portrait on the J10s. In contrast, with every other focussed feature test there were some counterfeits that produced error rates of 40% or higher. The integration of several security features into a single note is thus a necessary design strategy, to minimize the possibility that successful duplication of a single feature yields counterfeits that are mistaken for genuine. It is, however, equally important that currency handlers who authenticate notes be educated about the nature and use of the security features.

Analysis of the survey responses indicating degrees of reliance on the various security features generally did not predict how well participants performed in the actual experiment. Self-reports of performance, feature usage, and feature efficacy (as might be collected from surveys and focus groups) should, therefore, be interpreted with caution since such self reports may not reflect actual performance, feature usage or feature efficacy.

We believe that this study of the detection and perception of counterfeit banknotes with an experimental perspective has not only satisfied the BoC's objectives and provided a useful database for addressing future questions, it has revealed a rich arena for further research projects, two of which are described below.

1) Tracking eye movements and visual attention: Where do people look on a note, what feature do they look at first, then second, etc., for how long do they examine each note element and how many times are they going back to it if at all? We intend to answer these questions by monitoring eye position while observers look at banknotes with a range of design and security features. Converging evidence on the salience of note elements will also be obtained using a change or difference detection procedure in which observers look at two versions of the same note in which there is one changed element (cf, Shore & Klein, 2000).

2) Tactile inspection of notes: Using videotapes from the present study as data, we have conducted a detailed observational study (Gadbois et al., 2004) of note manipulation by CCH from Step 5's T and VT conditions. Studies examining the manipulation strategies of all three groups in Steps 1 and 3 would provide similarly valuable data.

In conclusion, the OSD was a standout performer when tested alone, and was undoubtedly partly responsible for the strong performance of the B20s. We urge its use, or the use of a similarly effective feature, in future series developed by the BoC. Conversely, the low level of intaglio was a significant contributor to the J10s poor performance. Future series should have sufficient intaglio to provide a salient degree of tactility. Because both of these features are resistant to optical duplication technologies, banknotes that incorporate them should provide counterfeit-discouragement for some time to come.

ACKNOWLEDGEMENTS

The research reported here was commissioned in October, 2002, by Linda Setlakwe on behalf of the BoC's Currency Development Team. We thank Josée Nadeau and the staff at the Halifax Branch of the BoC for graciously providing the space for us to conduct the research project. We appreciate the advice of Linda's colleagues in Ottawa, Gilbert Marois (Assistant Director of the Currency Development Team) and Dr. Sara Church on the research and this manuscript. Special thanks to Linda Setlakwe (not merely for asking Ray Klein if he would be interested in such a project, but more importantly) for her crucial roles as "note collector," project monitor, assiduous editor and enthusiastic collaborator.

REFERENCES

1. Berry, D. (1993). *Implicit learning: Reflections and prospects*. In A. Baddeley (Ed.), Attention: Selection, Awareness, and Control—A Tribute to Donald Broadbent (pp. 246-260). Oxford: Clarendon Press.
2. Collins, BL, Mayerson, SE & Worthey, JA (1985) *Noticeability of features of secure documents*. Technical report for the US Bureau of Engraving and Printing.
3. Croney, RGW (1970) *The relation between forgery and banknote design and the need to investigate the psychology of visual perception*. Paper presented to the Banknote Printer's Conference (Paris)
4. Croney, RGW (1974) *A summary of a research project on visual perception as applied to recognition of banknotes*. Paper presented to the Banknote Printer's Conference (Frankfurt)
5. Gadbois, S., Klein, RM. & Christie, J. (2004) *Tactile investigation of Canadian counterfeit and genuine banknotes*. Technical report for the Bank of Canada..
6. Herrigel, E (1953) *Zen in the Art of Archery (originally, Zen in der Kunst des Bogenschiessens,* 1948). translated by Richard F.C. Hull with an introduction by D.T. Suzuki. New York: Pantheon Books.
7. Hillstrom, AP & Bernstein, IH (2002) Counterfeit Detection for New and Old Currency Designs. in Rudolf L. van Renesse, Editor, Optical Security and Counterfeit Deterrence Techniques IV, *Proceedings of SPIE Vol. 4677*
8. Shore, DI & Klein, RM (2000) Change blindness: The effect of inversion. *Journal of General Psychology, 127*, 27-43.
9. Trowbridge, M. H. & Cason, H. (1932) An experimental study of Thorndike's theory of learning. *Journal of General Psychology*, *7*, 245-258.