THink: Inferring Cognitive Status from Subtle Behaviors

Randall Davis
MIT CSAIL
Cambridge, MA
davis@csail.mit.edu

David J. Libon
Neurology Department
Drexel College of Med.
david.libon@drexelmed.edu

Rhoda Au
Neurology Department
Boston Univ Sch of Med
rhodaau@bu.edu

David Pitman
MIT CSAIL
Cambridge, MA
dpitman@ mit.edu

Dana L. Penney
Department of Neurology
Lahey Health
dana.l.penney@lahey.org

Abstract

The Digital Clock Drawing Test is a fielded application that provides a major advance over existing neuropsychological testing technology. It captures and analyzes high precision information about both outcome and process, opening up the possibility of detecting subtle cognitive impairment even when test results appear superficially normal. We describe the design and development of the test, document the role of AI in its capabilities, and report on its use over the past seven years. We outline its potential implications for earlier detection and treatment of neurological disorders. We also set the work in the larger context of the THink project, which is exploring multiple approaches to determining cognitive status through the detection and analysis of subtle behaviors.

Introduction

We describe a new means of doing neurocognitive testing, enabled through the use of an off-the-shelf digitizing ballpoint pen from Anoto, Inc., combined with novel software we have created. The new approach improves efficiency, sharply reducing test processing time, and permits administration and analysis of the test to be done by medical office staff (rather than requiring time from clinicians). Where previous approaches to test analysis involve instructions phrased in qualitative terms, leaving room for differing interpretations, our analysis routines are embodied in code, reducing the chance for subjective judgments and measurement errors. The digital pen provides data two orders of magnitude more precise than pragmatically available previously, making it possible for our software to detect and measure new phenomena. Because the data provides timing information, our test measures elements of cognitive processing, as for example allowing us to calibrate the amount of effort patients are expending, independent of whether their results appear normal. This has interesting implications for detecting and treating impairment before it manifests clinically.

The Task

For more than 50 years clinicians have been giving the Clock Drawing Test, a deceptively simple, yet widely accepted cognitive screening test able to detect altered cognition in a wide range of neurological disorders, including dementias (e.g., Alzheimer's), stroke, Parkinson's, and others (Freedman 1994) (Grande 2013). The test instructs the subject to draw on a blank page a clock showing 10 minutes after 11 (called the "command" clock), then asks them to copy a pre-drawn clock showing that time (the "copy" clock). The two parts of the test are purposely designed to test differing aspects of cognition: the first challenges things like language and memory, while the second tests aspects of spatial planning and executive function (the ability to plan and organize).

As widely accepted as the test is, there are drawbacks, including variability in scoring and analysis, and reliance on either a clinician's subjective judgment of broad qualitative properties (Nair 2010) or the use of a labor-intensive evaluation system. One scoring technique calls for appraising the drawing by eye, giving it a 0-3 score, based on measures like whether the clock circle has "only minor distortion," whether the hour hand is "clearly shorter" than the minute hand, etc., without ever defining these criteria clearly (Nasreddine 2005). More complex scoring systems (e.g., (Nyborn 2013)) provide more information, but may require significant manual labor (e.g., use of rulers and protractors), and are as a result far too labor-intensive for routine use.

The test is used across a very wide range of ages – from the 20's to well into the 90's – and cognitive status, from healthy to severely impaired cognitively (e.g., Alzheimer's) and/or physically (e.g., tremor, Parkinson's). Clocks produced may appear normal (Fig. 1a) or be quite blatantly impaired, with elements that are distorted,

misplaced, repeated, or missing entirely (e.g., Fig 1b). As we explore below, clocks that look normal on paper may still have evidence of impairment.
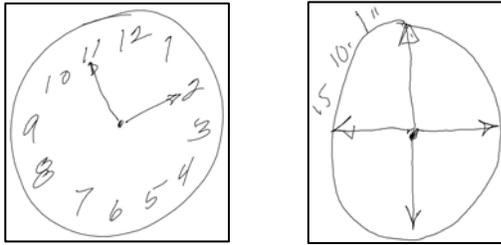

Figure. 1: Example clocks – normal appearing (1a) and clearly impaired (1b).

Our System

Since 2006 we have been administering the test using a digitizing ballpoint pen from Anoto, Inc.[1] The pen functions in the patient's hand as an ordinary ballpoint, but simultaneously measures its position on the page every 12ms with an accuracy of ±0.002". We refer to the combination of the pen data and our software as the Digital Clock Drawing Test (dCDT); it is one of several innovative tests being explored by the THink project.

Our software is device independent in the sense that it deals with time-stamped data and is agnostic about the device. We use the digitizing ballpoint because a fundamental premise of the clock drawing test is that it captures the subject's normal, spontaneous behavior. Our experience is that patients accept the digitizing pen as simply a (slightly fatter) ballpoint, unlike tablet-based tests about which patients sometimes express concern. Use of a tablet and stylus may also distort results by its different ergonomics and its novelty, particularly for older subjects or individuals in developing countries. While not inexpensive, the digitizing ballpoint is still more economical, smaller, and more portable than current handheld devices, and is easily shared by staff, facilitating use in remote and low income populations..

The dCDT software we developed is designed to be useful both for the practicing clinician and as a research tool. The program provides both traditional outcome measures (e.g., are all the numbers present, and in roughly the right places) and, as we discuss below, detects subtle behaviors that reveal novel cognitive processes underlying the performance.

Fig. 2 shows the system's interface. Basic information about the patient is entered in the left panel (anonymized here); folders on the right show how individual pen strokes have been classified (e.g., as a specific digit, hand, etc.).

Data from the pen arrives as a set of strokes, composed in turn of time-stamped coordinates; the center panel can show that data from either (or both) of the drawings, and is zoomable to permit extremely detailed examination of the data if needed.
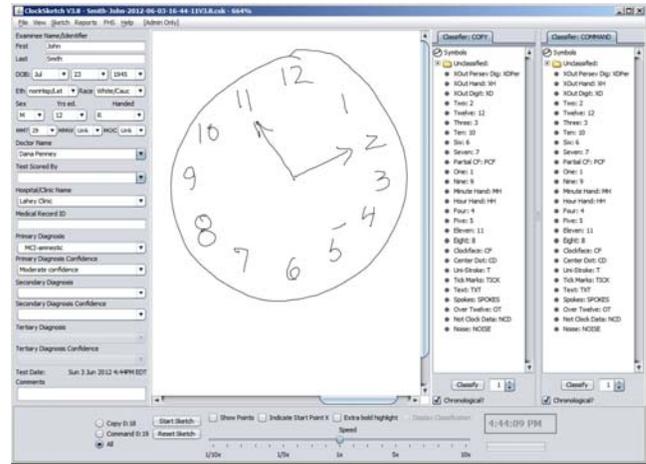

Figure 2: The program interface

In response to pressing one of the "classify" buttons, the system attempts to classify each pen stroke in a drawing. Colored overlays are added to the drawing (Fig. 3, a close-up view) to make clear the resulting classifications: tan bounding boxes are drawn around each digit, an orange line shows an ellipse fit to the clock circle stroke(s), green and purple highlights mark the hour and minute hands.
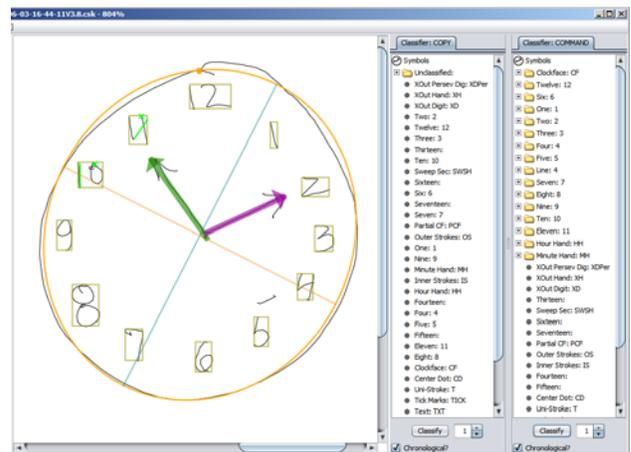

Figure 3: Partially classified clock

The classification of the clock in Fig. 2 is almost entirely correct; the lone error is the top of the 5, drawn sufficiently far from the base that the system missed it. The system has put the stroke into a folder labeled simply "Line." The error is easily corrected by dragging and dropping the errant stroke into the "Five" folder; the display immediately updates to show the revised classification. Given the system's initial attempt at classification, clocks

---

[1] Now available in consumer- oriented packaging from LiveScribe.

from healthy or only mildly impaired subjects can often be correctly classified in 1-2 minutes; unraveling the complexities in more challenging clocks can take additional time, with most completed within 5 minutes. The drag and drop character of the interface makes classifying strokes a task accessible to medical office staff, freeing up clinician time. The speedy updating of the interface in response to moving strokes provides a game-like feeling to the scoring that makes it a reasonably pleasant task.

Because the data is time-stamped, we capture both the end result (the drawing) and the behavior that produced it: every pause, hesitation, and time spent simply holding the pen and (presumably) thinking, are all recorded with 12 ms accuracy. Time-stamped data also makes possible a technically trivial but extremely useful capability: the program can play back a movie of the test, showing exactly how the patient drew the clock, reproducing stroke sequence, precise pen speed at every point, and every pause. This can be helpful diagnostically to the clinician, and can be viewed at any time, even long after the test was taken. Movie playback speed can be varied, permitting slowed motion for examining rapid pen strokes, or sped up for clocks by patients whose impairments produce vastly slowed motions. Movie playback can also be a useful aid when reviewing the classification of strokes in complex clocks.

The spatial and temporal precision of the data provides the basis for an unprecedented degree of analysis. The program uses the data to rapidly compute ~500 properties of the drawing that we believe are useful diagnostically (more on this below), including both traditional outcome measures and novel measures of behavior (e.g., pauses, pen speed, etc.). Because all the measurements are defined in software, they are carried out with no user bias, in real time, at no additional burden to the user.

The program makes it easy to view or analyze its results: it can format the results of its drawing analysis for a single test (i.e., the ~500 measurements) as a spreadsheet for easy review; provide comparison metrics on select variables of interest for glanceable clinician review; and can add the data to an electronic medical record. It also automatically creates two versions of each test result: One contains full patient identity data, for the clinician's private records, the second is de-identified and exportable to a central site where we accumulate multiple tests in a standard database format for research.

The program includes in the test file the raw data from the pen, providing the important ability to analyze data from tests given years ago with respect to newly created measures, i.e., measurements conceptualized long after that data has been collected.

The program also facilitates collection of data from standard psychometric tests (e.g., tests of memory,

intelligence), providing a customizable and user-friendly interface for entering data from 33 different user-selectable tests.

To facilitate the quality control process integral to many clinical and research settings, the program has a "review mode" that makes the process quick and easy. It automatically loads and zooms in on each clock in turn, and enables the reviewer to check the classifications with a few keystrokes. Clock review typically takes 30 seconds for an experienced reviewer.

The program has been in routine use as both a clinical tool and research vehicle in 7 venues (hospitals, clinics, and a research center) around the US, a group we refer to as the ClockSketch Consortium. The Consortium has together administered and classified more than 2600 tests, producing a database of 5200+ clocks (2 per test) with ground-truth labels on every pen stroke.

AI Technology Use and Payoff

As noted, our raw input is a set of strokes made up of time-stamped coordinates; classifying the strokes is a task-specific case of sketch interpretation. As with any signal interpretation task, we are dealing with plausible guesses rather than guarantees, especially given the range of inputs we must handle (e.g., Fig. 1b).

The program starts by attempting to identify subsets of strokes corresponding to three major elements: the clock circle, digits, and hands. The clock circle is typically the longest circular stroke (or strokes), often but not inevitably drawn first. The program identifies these and fits an ellipse to those points.

Our earliest approach to digit isolation and identification used the clock circle to define an annulus, then segmented the annulus into regions likely to contain each of the 12 numerals. Segmentation was done angularly, by a simple greedy algorithm: the angle to each stroke was measured from the estimated clock center, the angular differences ordered, and the (up to) 12 largest angular differences taken as segmentation points. Strokes in each segment were classified by the angular position of the segment (e.g., the segment near the 0-degree position was labeled as a "3").

Hand candidates are identified by finding lines with appropriate properties (e.g., having one end near the clock circle center, pointing toward the 11 or 2, etc.).

This extraordinarily simple approach worked surprisingly well for a wide range of normal and slightly impaired clocks.

We have since developed a far more sophisticated approach to digit isolation and identification able to deal with more heavily impaired clocks. It starts by using k-means to identify strokes likely to be digits, employing a

metric combining time (when it was drawn) and distance (stroke midpoint to clock center). The metric is based on the observation that numerals are likely to be further from the clock center and are usually drawn all at once. This set of strokes is divided into subsets likely to be individual digits using a novel representation we call spatio-temporal slices, that combines angle and timing information. In the simpler cases it performs much like the original angle-based segmenter described above, but offers the interesting ability to "unpeel" layers of ink resulting from crossed out and/or over-written digits, which often produce an incomprehensible collection of ink (e.g., Fig. 4).

Fig. 4: Overwritten digits (12 over-written with a 6)

The stroke subsets likely to be individual digits are then identified using a recognizer (Ouyang 2009) trained on digits from 600 clocks from healthy individuals. The recognizer works from visual features and is thus determining what the strokes look like, independent of how they were drawn (i.e., independent of order and timing).

Finally, context matters. The final classification of a set of strokes that could be either a 3 or a 5, for example, should of course depend in part on where in the clock face they appear. We accomplish this with a conditional random field trained on triples of angularly sequential digits from clocks drawn by healthy individuals, enabling the system to classify a digit based in part on the interpretation of the digits on either (angular) side of the current candidate.

The resulting system has >96% digit recognition on clocks from healthy individuals[2] along with the ability to unpack and interpret the otherwise incomprehensible layers of ink produced by crossing out and/or over-writing.

Machine Learning Results

The Clock Drawing Test is designed to asses diverse cognitive capabilities, including organization and planning, language and memory, and has utility for as a screening test, aiding in determining whether performance in any of these areas is sufficiently impaired as to motivate follow-up testing and examination. Our collection of several thousand classified clocks offered an opportunity to use machine learning to determine how much useful information there may be in the hundreds of features computed for each clock.

We selected three diagnoses of particular clinical interest and for which we had large enough samples: Alzheimer's

(n=146), other dementias (n=76) and Parkinson's (n=84). We used these diagnoses to explore the effectiveness of a large collection of machine learning algorithms, including SVMs, random forests, boosted decision trees, and others. They were trained to produce binary classifiers that compared each diagnosis against known-healthy patients (telling us whether the variables would be useful for screening), and each diagnosis against all conditions (telling us whether they would be useful in differential diagnosis).

In general, linear SVM's (Table I) produced the best results. As expected, the data sets are imbalanced. Accuracy rates are good and AUC's are acceptable, but the F1 scores are disappointing in some cases due to low precision rates. As the groups selected for this study have known clinical overlap (e.g., Parkinson's and other dementia may have comorbid Alzheimer's), low precision rates may (accurately) reflect this diagnostic overlap. We believe the classifiers may improve as we get additional tests from true positive subjects, and as we learn what additional features may be useful.

Parkinson's (P) vs Healthy (H)		
	class'd P	class'd H
P	56	28
H	52	732
Acc	0.908	AUC 0.74
F1	0.583	

Dementia (D) vs Healthy		
	class'd D	class'd H
D	48	28
H	84	476
Acc	0.824	AUC 0.70
F1	0.462	

Alzh. (Az) vs Healthy (H)		
	class'd Az	class'd H
Az	132	52
H	64	496
Acc	0.84	AUC 0.76
F1	0.69	

Parkinson's (P) vs All (-P)		
	class'd P	class'd -P
P	44	40
-P	204	1740
Acc	0.880	AUC 0.73
F1	0.265	

Dementia (D) vs All (-D)		
	class'd D	class'd -D
D	24	52
-D	304	1648
Acc	0.824	AUC 0.68
F1	0.119	

Alzheimer's (Az) vs All (-Az)		
	class'd Az	class'd -Az
Az	104	80
-Az	168	1676
Acc	0.878	AUC 0.65
F1	0.456	

Table I: SVM results[3]

One evident follow-up question is, how good is this result? What baseline do we compare it to? It would be best to compare to clinician error rate on judgments made from the same clock test data, but that is not available. There are, however, a number of established procedures designed to produce a numeric score for a clock (e.g., from 0-6), indicating where it sits on the impaired vs. healthy continuum. We are working to operationalize several of these, which as noted requires making computational something that was written for application by people, and that is often qualitative and vague. Once operationalized,

---

[2] Note that "drawn by healthy individuals" does not mean "free of error."

[3] The healthy counts differ in order to ensure age-matched comparisons.

we will be able to use these established procedures to produce their evaluation of all of our clocks. We will then train and test the performance of classifiers using each of those metrics as the test feature, and can then use this as a performance baseline against which to evaluate the results above.

We have also explored the use of a number of data mining algorithms, including Apriori, FPGrowth, and Bayesian List Machines (BLM) (Letham 2012), in an attempt to build decision models that balance accuracy and comprehensibility. Particularly with BLM, the goal is to produce a decision tree small enough and clear enough to be easily remembered and thus incorporated into a clinician's routine practice.

Clinical Payoff

Our dCDT program has had an impact in both research and practice. The collection and analysis of the wealth of high-precision data provided by the pen has produced insights about novel metrics, particularly those involving time, an advance over traditional clock drawing test evaluation systems, which focus solely on properties of the final drawing (e.g., presence/absence of hands, numbers, circle). The new metrics are valuable in daily practice and used by clinicians for their insight into the subject's cognitive status.

Time-dependent variables have, for example, proven to be important in detection of cognitive change. They can reveal when individuals are working harder, even though they are producing normal-appearing outputs. As one example, total time to draw the clock differentiates those with amnestic Mild Cognitive Impairment (aMCI) and Alzheimer's disease (AD) from healthy controls (HC) (Penney at al. 2014).

Additional pilot data suggests that AD can be distinguished from HC by comparing the percent of time spent during the task thinking (i.e., not making marks on paper, "think time") with the percent of time spent drawing ("ink time") independent of *what* was drawn (Penney et al. 2013b).

We have also defined a measure we call *pre-first-hand latency* (PFHL), measuring the precise amount of time the patient pauses between drawing the numbers on the clock and drawing the first clock hand. PFHL appears to distinguish normal patients from those with aMCI, AD, and vascular dementia (vAD) (Penney et al. 2011a).

These diagnostic groups also differed in how they spent their think time: PFHL and total clock drawing time are longer for AD and aMCI as compared to HC (Penney et al. 2011a). We believe PFHL is one measure of decision making, with longer latencies apparent in individuals with cognitive problems (like MCI). Importantly, our analysis of timing information means we detect these latencies even

in drawings that appear completely normal when viewed as a static image.

Other aspects of diagnostic significance overlooked by traditional feature-driven scoring include overall clock size, which we have found differentiates HC from aMCI (smaller), and aMCI from AD (smallest) independent of what is drawn. Patients with AD appear to work longer and harder, but produce less output (i.e., less ink and smaller clocks) when compared to cognitively intact participants (Penney et al. 2014).

Another novel variable we have uncovered concerns the seemingly inadvertently produced ink marks in a clock drawing: these are pen strokes contained within the clock face circle, but not clearly identifiable as clock elements. One class of strokes that we call "noise" has typically been thought to be meaningless and is ignored in traditional analysis. Yet one of our studies (Penney et al. 2011b) found interesting information in the length and location of these strokes. We found that healthy controls (HC) made very few of the smallest noise strokes (those <0.3mm), while clinical groups, including those with MCI, made significantly more longer noise strokes, distributed largely in the upper right and left quadrants of the clock (i.e., locations on the clock where patients must negotiate the request to set clock hands to read "10 after 11"). We hypothesize that noise strokes may represent novel hovering-type marks associated with decision-making difficulty in time-setting.

All of these features are detected and quantified by the program, producing information that clinicians find useful in practice.

Development and Deployment

From its inception this project has been produced with quite spare resources (i.e., a few small seedling grants). The Java code base has (and continues to be) produced by a sequence of talented undergraduate programmers and one computer science professor, inspired, guided, and informed by a few highly experienced neuroscientists who donate their time because they see the opportunity to create a fundamentally new tool for assessing cognitive state.

We started by reconceptualizing the nature of cognitive testing, moving away from the traditional approach of "one test one cognitive domain" (e.g., separate tests for memory, executive function, etc.) with standard outcome measures, re-focusing instead on the cognitive processes inherent in the drawing task. We broke the complex behavior of clock drawing down to its most basic components (pen strokes), but ensured that we also captured behavioral aspects. We piloted the program at a key clinical site, collecting hundreds of clocks that were used to further refine our set of measurements.

We developed a training program for technicians who administer the test and classify strokes, and adapted the software to both PCs and Macs. We recruited beta testing sites throughout the US, forming the ClockSketch Consortium. We hosted user training sessions to ensure standard testing and scoring procedures, and measured user proficiency across sites. We established a collaboration with the Framingham Heart Study, a large scale epidemiological study, to enable the development of population based norms for our measures.

Our ongoing development was significantly aided by a key design decision noted above – the raw pen data is always preserved in the test file. This has permitted years of continual growth and change in the measurements as we discover more about the precursors to cognitive change, with little to no legacy code overhead.

Our deployment strategy has been one of continual refinement, with new versions of the system appearing roughly every six months, in response to our small but vocal user community, which supplied numerous suggestions about missing functionality and improvements in the user interface.

One standard difficulty faced in biomedical applications is approval by internal review boards, who ensure patient safety and quality of care. Here again the use of the digital pen proved to be a good choice: approval at all sites was facilitated by the fact that it functions in the patient's hand as an ordinary pen and presents no additional risk over those encountered in everyday writing tasks.

Next Steps

The Digital Clock Drawing Test is the first of what we intend to be a collection of novel, quickly administered neuropsychological tests in the THink project. Our next development is a digital maze test designed to measure graphomotor aspects of executive function, processing speed, spatial reasoning and memory. We believe that use of the digital pen here will provide a substantial body of revealing information, including measures of changes in behavior when approaching decision points (indicating advanced planning), length of pauses at decision points (a measure of decision-making difficulty), changes in these behaviors as a consequence of priming (a measure of memory function), speed in drawing each leg of a solution (measures of learning/memory), and many others.

Capturing these phenomena requires designing mazes with new geometric properties. Because maze completion is a complex task involving the interplay of higher-order cognition (e.g., spatial planning, memory), motor operations (pen movement) and visual scanning (eye movement to explore possible paths), little would be gained by simply using a digital pen on a traditional maze.

We have designed mazes that will distinguish the phenomena of interest.

We have also created mazes of graded difficulty, accomplished by varying characteristics of the maze, as for example the number of decision points and the presence of embedded choice points. These features will allow us to explore difficulty-tiered decision making by measuring changes in speed approaching a decision point, the length of pauses at each of those points, and by detecting and analyzing errors (e.g., back-tracking, repetition of a wrong choice, etc). Tiered decision making is in turn an important measure of executive function that will enable us to detect, measure, and track subtle cognitive difficulty even in correctly solved mazes. We hypothesize that individuals with subtle cognitive impairment, as in MCI and other insidious onset neurologic illness (e.g., AD, PD), will pause longer than healthy controls at more difficult decision junctions, while demonstrating only brief or no pauses at easier junctions. We posit that these pauses will be diagnostic even when the correct path solution is chosen. The inclusion of tiered difficulty will allow us to grade cognitive change by assessing cognition at various levels of decision-making difficulty.

The subject will be asked to solve two mazes in sequence, both of which (unknown to the subject) are identical, except that the first has no choice points (added walls remove all choices). This in turn will permit calibrating the effects of priming, giving an indication of the status of memory. The comparison of these two tasks – with identical motor demands and identical solutions – enables using the subject as their own control, and using difference scores from the first to second maze will help parse out potential confounding factors (e.g., fatigue, depression). These ideas are just the beginning of what appears to be possible with an appropriately designed maze and the data made available with the digital pen.

Larger Implications

Insights about Cognition

One interesting consequence of the detailed data we have is the light it may shed on some previously unknown (or at least under-appreciated) behavioral phenomena that opens up a new approach to understanding cognition. One of these is a phenomenon we call "hooklets." Fig. 5 below shows a zoomed-in view of an 11, showing that there is a hook at the bottom of the first "1" that heads off in the direction of the beginning of the next stroke. While sometimes visible on paper, hooklets are often less than 0.5mm long, not visible on the paper, yet are clear in the digital record and are detected automatically by our program.

We have hypothesized (Lamar 2011) (Penney et al. 2013a) that hooklets represent anticipation: the subject is thinking about the next stroke and begins moving in that direction before finishing the current one. This is revealing, as the ability to think ahead is a sign of cognitive health: impaired cognition can limit capacity to multi-task, leaving resources sufficient only to attend to the current moment. If hooklets are indeed a sign of cognitive health, we have the intriguing possibility that their progressive *disappearance* may be a (perhaps early) sign of cognitive decline, as in pre-clinical Alzheimer's.

Figure 5: A hooklet

By focusing on the component processes of cognitive function applied to a standard task and moving away from a traditional approach based on outcome error, we open up the opportunity to study the subtle changes in cognitive health that herald cognitive change *before problems manifest*. Understanding cognitive strategies that emerge when individuals are consciously or unconsciously compensating for emerging impairment may enable the detection (and hence treatment) of medical conditions far earlier than currently possible, as well as assist with developing new treatments and monitoring their efficacy. Potential implications of earlier detection and treatment are evident even when applied to just one disease, Alzheimer's dementia. An estimated 5.4 million Americans had Alzheimer's disease in 2012, this number is estimated to rise to 6.7 million by 2030 and projected to reach 11-16 million by 2050 if no medical developments alter the disease process. Healthcare costs for 2012 are estimated at $200 billion and are projected to rise to a staggering $1.1 trillion in 2050 (Alzheimer's 2012). The additional human costs of care-giving, including lost quality of life and suffering, are immeasurable. Early detection of this disease at a stage that afforded intervention while it was still preclinical or presymptomatic could thus result in substantial benefits.

Insights about Assessment

Current practice in cognitive assessment typically (and unsurprisingly) assumes that average test scores indicate absence of impairment. We suggest otherwise. We believe that patients often unwittingly hide early, and thus subtle, impairment behind compensatory strategies, for example thinking harder or working longer in ways that are typically not visible to an observer. Their final results may appear normal (e.g., a clock drawing that looks normal),

but an ability to "see through" compensatory strategies would detect the additional mental work and the brief but important additional time spent on a task.

We hypothesize that this can be done by detecting and measuring extremely subtle behaviors produced without conscious effort, as, for example the brief, inadvertent pauses in a task, or the seemingly accidental pen strokes (both noted above), that are normally overlooked or considered spurious and ignored. Detecting and measuring these subtle behaviors reveals the effort normally camouflaged by compensatory strategies (Penney et al. 2014).

We believe this approach to assessment will make possible considerably more detailed information about the cognitive status of an individual, with significant implications for diagnosis and treatment.

Related Work

Over its long history numerous scoring systems have been proposed for the CDT (see, e.g., (Strauss 2006)), but as noted above they may present difficulties by relying on vaguely worded scoring criteria (producing concerns about reliability) or by requiring labor-intensive measurements.

Recent work on automating the clock drawing test is reported in (Kim 2013), where it was administered on and analyzed on a tablet. That work focused on interface design, seeking to ensure that the system was usable by both subjects and clinicians. It makes some basic use of the timing information available from the tablet, does basic digit recognition and some analysis of patient performance, but is unclear on how much of the patient performance analysis was done by the system vs. by the clinician. It does not report dealing with the complexities of the sort noted above, like over-written and crossed out digits, and appears reliant on traditional scoring metrics..

Work in (Sonntag 2013) reports on another medical application of the Anoto pen, employing it to annotate medical documents in ways well suited to a pen-based interaction (e.g., free-form sketching). The resulting system offers the ease and familiarity of recording information by writing, with the added ability to analyze the annotations and hence integrate them into the medical record.

Work in (Tiplady 2003) reported using the Anoto pen in attempting to calibrate the effects of ethanol on motor control (i.e., detecting impaired drivers), by having subjects draw small squares as quickly as possible, a very limited experiment but one that attempted to use data from the pen to detect impaired behavior.

Adapx Inc. developed prototype pen-enabled versions of several standard neuropsychological tests, including trail-making, symbol-digit, and Reys-Osterreith complex figure.

Each of these demonstrated the ability to collect digitized data, but did not do data analysis (Salzman 2010).

Summary

The Digital Clock Drawing Test has demonstrated how the original conception and spirit of the clock drawing test can be brought into the digital world, preserving the value and diagnostic information of the original, while simultaneously opening up remarkable new avenues of exploration. We believe the work reported here takes an important step toward a new approach to cognitive assessment, founded on the realization that we no longer have to wait until people look impaired to detect genuine impairment. This offers enormous promise for early differential diagnosis, with clear consequences for both research and treatment.

Acknowledgements

References

Alzheimer's Association, 2012 Alzheimer's Disease Facts and Figures, *Alzheimer's & Dementia*, Volume 8, Issue 2, 2012.

Freedman, M., Leach, L., Kaplan, E., Winocur, G., Shulman, K., & Delis, D. 1994. *Clock Drawing: A Neuro-psychological Analysis*, Oxford Univ Press.

Fryer S., Sutton E., Tiplady B., Wright P. 2000. Trail making without trails: the use of a pen computer task for assessing brain damage. *Clinical Neuropsychology Assessment* 2, 151-165.

Grande, L., Rudolph, J., Davis, R., Penney, D., Price, C., Swenson, R., Libon, D., & Milberg, W. 2013. Clock Drawing: Standing the Test of Time, in *The Boston Process Approach to Neuropsychological Assessment*, Ashendorf, Swenson, Libon (eds.), Oxford Univ Press 2013.

Kim H, 2013. The Clockme System: Computer-Assisted Screening Tool For Dementia, PhD, Thesis, Georgia Institute of Technology, College of Computing, 2013.

Lamar, M., Grajewski, M.L., Penney, D.L., Davis, R., Libon, D.J., and Kumar, A. 2011. The impact of vascular risk and depression on executive planning and production during graphomotor output across the lifespan. Paper at *5th Congress of the Int'l Soc. for Vascular, Cognitive and Behavioural Disorders*, Lille, France.

Letham B, Rudin C, McCormick T, Madigan D, 2012. Building interpretable classifiers with rules using Bayesian analysis, Univ Washington Dept of Statistics Technical Report, TR609.

Nair, A. K., Gavett, B. E., Damman, M., Dekker, W., Green, R.C., Mandel, A., Auerbach, S., Steinberg, E., Hubbard, E. J., Jefferson, A., &

Stern, R. A. 2010. Clock drawing test ratings by dementia specialists: interrater reliability and diagnostic accuracy. *J Neuropsychiatry Clin Neurosci*. 2010 Winter; 22(1):85-92.

Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H. 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment, *J Am Geriatr Soc*. 2005 Apr;53(4):695-9.

Nyborn, J. A., Himali, J. J., Beiser, A. S., Devine, S. A., Du, Y., Kaplan, E., O'Connor, M. K., Rinn, W. E., Denison, H. S., Seshadri, S., Wolf, P. A., & Au, R. 2013. The Framingham Heart Study clock drawing performance: normative data from the offspring cohort. *Clock scoring: Exp Aging Res*. 2013;39(1):80-108.

Ouyang T, Davis R. 2009. A visual approach to sketched symbol recognition, *Proc IJCAI 2009, pp 1463-1468*.

Penney, D.L., D. J. Libon, M. Lamar, C. C. Price, R. Swenson, S. Scala, J. Eppig, C. Nieves,C., K. D.Garrett, R. Davis. 2011a. The Digital Clock Drawing Test (dCDT) - IV: Clock Drawing Time and Hand Placement Latencies in Mild Cognitive Impairment and Dementia. Abstract and poster, *5th Congress of the Int'l Soc. for Vascular, Cognitive and Behavioural Disorders*, Lille, France.

Penney, D.L., Libon, D. J., Lamar, M., Price, C. C., Swenson, R., Eppig, J., Nieves, C., Garrett, K. D., & Davis, R. 2011b. The Digital Clock Drawing Test (dCDT) - I: Information Contained Within the "Noise". Abstract and poster, *5th Congress of the Int'l Soc. for Vascular, Cognitive and Behavioural Disorders*, Lille, France.

Penney, D. L., Lamar, M., Libon, D. J., Price, C. C., Swenson, R., Scala, S., Eppig, J., Nieves, C., Macaulay, C., Garrett, K. D., Au, R., Devine, S., Delano-Wood, L., & Davis, R. 2013a. The digital Clock Drawing Test (dCDT) Hooklets: A Novel Graphomotor Measure of Executive Function. Abstract and poster, *6th Congress of the Int'l Soc. for Vascular, Cognitive and Behavioural Disorders*, Montreal, Canada.

Penney, D.L, Libon,D.J., Lamar, M., Price, C. C.,Swenson, R., Scala, S., Eppig, J., Nieves, C., Macaulay, C., Garrett, K. D., Au, R., Devine, S., Delano-Wood, L., and Davis, R. 2013b. The digital Clock Drawing Test (dCDT) In Mild Cognitive Impairment and Dementia: It's a Matter of Time. Abstract and poster, *6th Congress of the Int'l Soc. for Vascular, Cognitive and Behavioural Disorders*, Montreal, Canada.

Penney, D.L., Libon, D.J., Au, R., Lamar, M., Price, C.C Swenson, R., Macaulay, C., Garrett, K.D., Devine, S., Delano-Wood, L., Scala, S., Flanagan, A., and Davis, R. 2014. Working Harder But Producing Less: The Digital Clock Drawing Test (dCDT) Differentiates Amnestic Mild Cognitive Impairment And Alzheimer's Disease. Abstract and Poster, *42nd Meeting of the Intern'l Neuropsychological Society*. Seattle, Washington. 2014.

Salzman K, Cohen P, Barthelmess P, 2010. ADAPX Digital Pen: TBI Cognitive Forms Assessment. Report to US Army Medical Research and Materiel Command, Fort Detrick, MD, July 2010.

Sonntag D, Weber M, Hammon M, Cavallo A, 2013. Integrating digital pens in breast imaging for instant knowledge acquisition, *Proc IAAI 20013,* pp. 1465-1470.

Strauss E, Sherman E, Spreen O, 2006. *A Compendium of Neuropsychological Tests*, pp. 972-983, Oxford University Press, 2006.

Tiplady B, Baird R, Lutcke H, Drummond G, Wright P. 2013. Use of a digital pen to administer a psychomotor test, *Journal Of Psychopharmacology*. 17(supplement 3): A71.