Clinical and Legal Decision Making: How Education May Affect Cognitive Reasoning Fallacies

Jonathan G. Lash

Cornell University

A Thesis Presented to the Faculty of the Graduate School of Cornell University in Partial Fulfillment of the Requirements for the Degree of Master of Arts

by Jonathan G. Lash

January 2017

Abstract

Previous research suggests that there are differences in the way various professionals reason. This is evidenced by variable susceptibility to cognitive reasoning fallacies. This study examines these differences between lawyers and doctors through a battery of cognitive reasoning measures including the CRT, Numeracy Scale, Linda Problem, Wason Selection Task, and Beads Task. Secondly, this study aims to pinpoint when these differences emerge in each respective professional track—prior to the onset of training vs. later during professional practice--as a way to discern a self-selecting bias when it comes to thinking styles and career choices. We found that while not ubiquitous, there was evidence to suggest both loci. For certain cognitive aspects in legal education it seems that more schooling leads to improved performance in various forms of reasoning. However, this did not hold true for medical education, which seemed to have relatively little effect on performance. Generally, lawyers were able to score higher than doctors on our battery of cognitive tests, which suggests either that medical professionals are not able to make decisions as soundly, or that this study was unable to reveal the areas of cognitive functioning in which doctors excel over lawyers. Results from this study should be extended by future research that examines more closely the differences revealed here, to better understand how professional education affects decision making.

Acknowledgements

I'd like to first thank Dr. Stephen Ceci and Dr. Wendy Williams for their help and support throughout my time at Cornell. I'd like to thank Kayla Burd for her guidance and generosity in helping me complete this project. I am also grateful for the love and support of my family and friends.

Biographical Sketch

Jon Lash graduated from Cornell University with a Bachelor of Science in 2016. He is pursuing a Masters degree in Human Development at Cornell and is expected to graduate January 2017. He plans to attend medical school Fall of 2017 to earn his Medical Doctorate.

Clinical and Legal Decision Making: How Education May Affect Cognitive Reasoning Fallacies

Introduction

Choosing a career is one of the most important decisions a person makes over his or her life. It determines how you spend the majority of your adult life and it becomes a large part of your self identity. Other people will even use your job as a way to assess your role in society: how much money you make, or even how smart you are. It is interesting then to think of some of the traditional advice given to people at such a critical juncture. We often hear advice that falls into one of two categories: the idealistic or the pragmatic. Concerning the former, some suggest we follow our dreams and that if we work hard enough we can attain anything. Then there are others who suggest that we act in a more measured way—that we should develop a skill and find a high-paying profession that we count on to support ourselves and our family. Although these two perspectives aren't mutually exclusive, they can be in conflict, and for those on the brink of choosing a career, confusing. Ultimately all any of us can do is make the best decision we can with what we know at the time.

Unfortunately, this decision-making process is often flawed. Human reasoning is inherently based on our ability to process information, for which everyone has a limit. Therefore, we must rely on gestalt principles or heuristics. These general ideas, similar to things such as "rules of thumb" or "educated guesses" can be especially useful when our cognitive resources are taxed, such as when we are stressed or need to make snap decisions (Neth & Gigerenzer, 2015). Heuristics are often accurate and efficient bases of decision-making, but can sometimes include unrelated and extraneous information that can lead to misjudgments. Just as a rule of thumb cannot apply always, neither can heuristics be accurate for all situations. This makes sense considering that our experience tells us that many people, when presented with the same information, will often come to disparate conclusions. The failure of our decision-making process has been shown to be due, at least in part, to the incorporation of bits and pieces of stored information that may only relate tangentially to the situation at hand. These manifest themselves in the form of biases in our decision-making process (Gigerenzer & Gaissmaier, 2011). One of the largest sources of variation in decision-making research stems from individual differences (Hofmans et al., 2013), and it makes sense that these differences arise from differences in the way certain biases get erroneously incorporated into the heuristics we often rely on.

Does this mean that, when presented with the facts, we are somewhat doomed to form biased conclusions and misjudgments? No, but it does mean that people are susceptible to cognitive errors and fallacies. Interestingly though, there is evidence to suggest that people in different professions make decisions differently. Further, this decision-making bias seems to be somewhat consistent among people in that profession. For instance, Hoffrage and Gigerenzer showed that physicians, when presented with a question regarding the percentage in which a diagnostic test would yield positive results, only correctly estimated 10% of the time. This was even after information was provided such as the base rate of the disease in the population, the type 1 error rate of the test, and the correct positive result rate of the test. These findings along with the results of Berwick et al. reflect negatively on doctors' ability to reason statistically, but may hint at a common cognitive bias. Loftus et al. found that people in creative careers such as artists and architects showed superior memory when compared to law enforcement agents. These underlying, profession-specific, cognitive deficits beg the question: is there selection bias going on such that people with a certain cognitive profile choose one career disproportionately, or does the nature of the career shape the way its professionals think, independently of any preexisting characteristics?

This study sought to shed some light on this question by attempting to examine the effect of educational training on the decision-making process--specifically, how progress along an education track (e.g. medicine, law, engineering etc.) changes the frequency and type of errors made. To do this, students and graduated professionals at all stages of their career, from pre-professional undergraduates to graduate students to newly-minted professionals to seasoned veterans in the field were given a battery of critical thinking measures via an online survey. We hypothesized that across all professions, ability to reason, and thus avoid cognitive pitfalls, would peak during last years of schooling since this is the epitome of the emphasis on formal cognition. Additionally, we hypothesized that doctors and lawyers would, on average, outperform other members of the general population on the battery of tests administered based on their generally higher cognitive test performance. We hypothesized that doctors would prefer to make their decisions based on a smaller amount of information, and finally, we hypothesized that reasoning differences (i.e. a specific profile formed from performance on the battery) would emerge between professions and become more pronounced as years spent in that profession increased.

Methodology

Participants

A total of 736 participants completed all tasks included in the present study. Two-hundred-thirty-nine participants were recruited via Cornell University's Psychology Experiment System (SONA), and 203 participants were recruited from Amazon's Mechanical Turk (MTurk). Additionally, 294 individuals were recruited via snowball sampling. Approximately 250 individuals started the survey, but failed to complete all measures. Their data was included for the measures they did complete.

Participants were offered incentives based on their recruitment platform: participants recruited via SONA were awarded 1 SONA credit for their participation, while those who were recruited from Amazon's MTurk were given $0.50 for their participation. Individuals recruited by snowball sampling (e.g. via email) were not compensated, but instead thanked for their participation. Of the entire sample, 28.5% participants identified as male, 45.6% as female, 0.3% did not identify with either gender, and 25.5% of participants, did not specify their gender. Participant ages ranged from 18 to 78 (M = 29.75 SD = 13.6).

Design

This study employed a cross-sectional, correlational design with an exploratory component that compared individuals' performances on a battery of tasks. These tasks included measures of critical thinking and decision-making. Individuals were grouped according to self-reported professional field, age, gender, income, or ethnicity. However, regardless of grouping, all participants were shown an identical battery of tasks. Each participant's scores on cognitive measures were then used in subsequent analyses.

Measures

Professional field. Participants first completed a questionnaire regarding their intended or current professional track. If the participant indicated that he or she was enrolled as a student in either medical school or law school, they were asked if they intended on specializing in a specific field of law or medicine. Participants that indicated they received their JD or MD degree, were asked to include their specialty, if applicable, and the number of years since receiving his or her degree. Additionally, participants who indicated that they were neither interested in medicine nor law were asked to specify their level of education (Some college, Bachelor degree, or graduate degree) as well as their field of study and years since receiving their degree if they had done so.

Cognitive Reflection Test (CRT). The Cognitive Reflection Test (Frederick, 2005) is a short (3 question) questionnaire that measures cognitive reflection. The questions included in the test have obvious, but incorrect answers. They require the participant to override their “gut” reaction and engage in further consideration to find the right answer. These different types of thinking are called “System 1” and “System 2” processes by Stanovich and West (2000). System 1 processing is a quick gut reaction that is often below conscious awareness and does not tap limited-capacity attentional resources, whereas System 2 processing is a slower and more reflective thought process. The CRT is used to assess a participant’s ability to override system 1 thinking and engage in system 2 thinking. The ability to do so is highly correlated with patience and propensity to delay gratification (Frederick, 2005). For example, an individual scoring high on cognitive reflective abilities would be able to appreciate a larger reward that was presented days or weeks away compared to a smaller immediate reward. Interestingly (Mischel, 1974), and Shoda, Mischel and Peake (1990) showed that increased propensity for a delayed reward in preschool predicted higher SAT scores over a decade later. Additionally, CRT scores can be used to assess how an individual's decision-making process employs heuristics. It can also be used to indirectly assess how quickly someone forms decisions. The inability to re-think an initial gut reaction during processing may indicate a preference for quicker decisions. Liberali, Reyna, Furlan, Stein, and Pardo (2012).

Abridged numeracy scale- The abridged numeracy scale used is part of a longer, 11-item scale developed by Lipkus, Samsa, and Rimer (2001). Six of the original 11 questions were eliminated for the task because they had exceptionally high proportions of correct responses (over 75% correct). We felt they wouldn't add any predictive power to the task. An example of a question used in the task was: "*Imagine that we rolled a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up even (2, 4, or 6)?*" These questions were used to gauge participants' ability to comprehend and manipulate numeric data. Research has shown that this skill may have an important impact on decision-making (Estrada, Barnes, Collins, & Byrd, 1999; Reyna, Nelson, Han, & Dieckmann, 2009). The score on the Numeracy task is related to the CRT because an exceptionally low score on this task could impede a participant's ability to complete the CRT successfully regardless of their actual cognitive reflection ability.

The Linda Conjunction Fallacy problem. Participants' ability to make rational judgments in a probabilistic context was tested with the Linda Problem (Tversky & Kahneman, 1983). In the Linda problem a woman named Wendy (usually Linda) is described as a socially liberal woman invested in "*issues of discrimination and social justice*". Participants are then asked whether Wendy would be more likely to be a "*bank teller*" or "*a bank teller and active in the feminist movement*". When answering the Linda problem, many participants find themselves making a conjunction fallacy. This is a type of fallacy that occurs when it is assumed that specific conditions are more probable than a single general one. It is always true that the likelihood of two events happening simultaneously is less than or equal to the likelihood of either one occurring alone. This happens when participants rely too heavily on irrelevant information; in this case, the part of the problem that portrays Wendy as someone committed social reform. Previous research suggests that an individual's performance on the Linda problem may be, similar to performance on the CRT, related to over-reliance on heuristic based decision-making (Tversky & Kahneman, 1983).

Wason selection task. The Wason Selection Task (Wason, 1968) measures participants' deductive reasoning. First, participants were asked to answer an "if–then" logic puzzle using four white-faced cards labeled with either black-type letters or numbers (Specifically: "A", "K", "4", and "7"). Participants were told that each card had a number on one side, and a letter on the other. Participants were presented with the following statement: "*If a card has an even number on one side, then it has a vowel on the other*". They were then asked to chose two of the four cards to "flip over" in order to verify that the rule above was true. In other words, which two cards' backside could contain information capable of disproving the statement given. For example, if the card with the number "4" on the front was flipped, and shown to have the letter "R" on the backside, the rule would be disproven. However, if the card with the number "7" was flipped, neither a vowel nor a consonant on the backside could disprove the statement because the rule only applies to cards with and even number on one side.

Beads task. The tendency for individuals to "jump to conclusions" was tested by the Beads Task (Garety et al., 1991, Garety & Freeman, 1999). The Jumping to conclusions beads task, as defined by Garety (1991), is a measure of a person's tendency to gather information before making a decision, and some even characterize low information gathering (a high tendency to jump to conclusions) as a type of decision-making bias (So et al. 2015). Hasty decision-making can reasonably lead to flawed decisions when insufficient information is gathered before coming to a conclusion. Although previous research has used the Beads Task to assess this bias in delusion-prone individuals (Warman, Lysaker, Martin, Davis, & Haudenschield, 2006), some new research has attempted to use the Beads Task on healthy, neurotypical individuals (So et al. 2015). In this study, the beads task was used to assess the nature of information gathering in decision-making for different professional groups.

In the present study, two versions (Easy and Hard) of the Beads Task were administered to participants. In the easy version of the Beads Task, participants were presented with pictures of two jars (labeled "A" and "B", respectively). Each Jar contained 100 beads, and in the Easy version, Jar "A" was comprised of 15% blue beads and 85% red beads. Jar B had a ratio of 85:15 Blue beads to Red beads. Next, participants were told "*The computer will now select one of the jars and draw random beads from it*". Participants were allowed to continue to draw beads from the same jar until they felt comfortable that they knew which jar the computer had selected. Each participant could, at each bead draw, decide to make a guess or draw another bead. The same set of instructions was given during the Hard Beads Task. The ratio of blue to red beads in the Hard condition was 60:40, making the choice of jar less obvious and often requiring participants to draw more beads. The bead sequence for both conditions are pictured below.

Easy 85:15 beads task (○ = Blue, ● = Red):

○ ● ● ● ● ○ ● ● ● ● ● ○ ● ● ○ ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●

Hard 60:40 beads task (○ = Blue, ● = Red):

○ ○ ● ● ● ○ ○ ○ ● ○ ○ ○ ○ ○ ○ ● ○ ○ ● ○ ● ● ○ ○ ● ● ○ ○ ● ○ ○

Demographic questionnaire. Lastly, participants were given a battery of demographic questions, including questions regarding age; gender; race; highest level of education completed; number of academic courses taken in STEM fields; and estimated income. For those participants still dependent on a parent or guardian's income, he or she was asked to estimate the income of that parent or guardian.

Attention checks. All participants were timed while completing the survey. Participants who spent less than 70 seconds on the survey were excluded from subsequent analyses.

Procedure

Participants completed the survey online through the Qualtrics survey tool. Participants were given access to the study based on how they were recruited. Those recruited through SONA systems received the link there while others were given access through personal emails or Amazon's MTurk.

The link lead participants to the Cornell sponsored Qualtrics website where they were able to continue. Each gave informed consent for their participation. The first part of the survey consisted of the participants intended or current professional track. Then, participants completed the battery of twelve cognitive and decision-making tasks (detailed above) in random order. specifically, participants answered three CRT questions, five Abridged Numeracy Scale questions, the Linda problem, the Wason Selection Task, and two trials (one easy, and one hard) of the Beads Task. Lastly, participants completed demographic questions.

Data Analysis and Results

Cognitive Reasoning Task-

A One-Way ANOVA test was performed to compare mean values for a composite CRT score. The data was blocked into ten comparison groups including: Pre-Medical students, Pre-Law Students, Medical Students, Law Students, Doctors, Lawyers, Psychologists, Engineers, Heath Support Staff, and a group labeled "other" that included the rest of the general population. The omnibus model was significant with $F(7, 718) = 6.25$, $p < .001$, $\eta p^2 = .057$. Post hoc pairwise comparisons were made using Bonferroni corrections and this revealed several group differences. After Bonferroni correction, pre-medical undergraduates had significantly lower CRT scores than Medical school students ($p=0.024$), law students ($p=0.003$), lawyers ($p=0.01$), and engineers ($p=0.017$) (See Table 1 and Figure 1).

Table 1- CRT Score- Between Group Comparisons

Educational Track	n	Mean	Standard Deviation
Pre-Med	116	1.51	1.18
Medical Student	43	2.16	.97
Medical Doctor	53	1.79	.91
Pre-Law	54	1.74	1.09
Law Student	60	2.18	0.98
Lawyer	45	2.30	0.90
Engineer	73	2.12	1.16
Other	282	1.63	1.07
Total	726	1.80	1.12

Within group comparisons

When comparing within groups, we used the same one-way ANOVA procedure with Bonferroni post-hoc correction for pairwise comparisons. The purpose of these within group comparison was to better elucidate the source of differences seen on performance on these tasks. We were interested in examining whether formal education changes performance on a task, or if there were pre-existing differences in the population that are unchanged by education.

For the law track our model $F(2, 158) = 4.00$, $p = .02$, $\eta p^2 = .048$, found that Pre-Law students had a lower CRT score than law students ($p=0.031$), but interestingly, they were not significantly different from practicing lawyers (See table 2).

Table 2- CRT Score- Within group Comparison for Law track

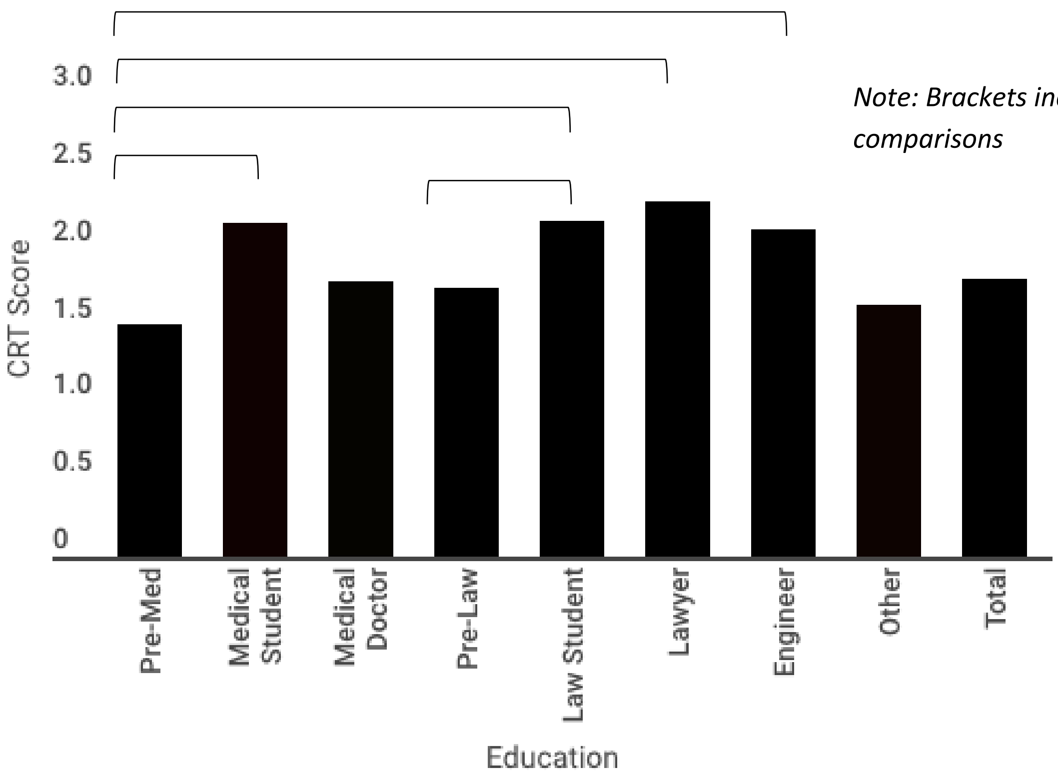
Educational Track	n	Mean	Standard Deviation
Pre-Law	54	1.74	1.09
Law Student	60	2.18	0.98
Lawyer	45	2.30	0.90

This was the same trend seen in pre-medical students who were outperformed by medical students ($p=.002$) but did not differ from practicing doctors $F(2, 211) = 6.34$, $p = .002$, $\eta p2 = 0.057$. No other comparison's attained significance (see table 3). These two findings showing that students prior to professional training fared worse on this task than medical or law school students, but they did not fare worse than actual physicians and attorneys, suggests that sampling may be an issue, or else some real-life experiences of doctors and attorneys work to shift them to more impulsive decision-making of the sort observed prior to the onset of professional training.

Table 5- CRT Score- Within group Comparison for Medical track

Educational Track	n	Mean	Standard Deviation
Pre-Med	116	1.51	1.18
Medical Student	43	2.16	0.97
Doctor	53	1.79	0.91

Figure 1- Mean CRT score comparison



Note: Brackets indicate significant comparisons

Abridged Numeracy Scale

The numeracy scale here was used to assess participants' ability to understand basic probability and mathematical concepts. Nisbett and Lehman (1990) showed that there were significant differences in statistical reasoning ability between graduate students in different disciplines. In order to expand on this result through a within group comparison we performed a one-way ANOVA to compare mean composite scores on the Numeracy Scale. The omnibus model was significant and revealed an effect of education on mean score, $F(7, 712) = 3.12$, $p = .003$, $\eta p^2 = .03$. Post-hoc pair wise comparisons again utilizing the Bonferroni correction revealed that pre-med students performed significantly worse than medical school students (p=0.021). No other comparisons yielded significant differences (See table 4).

Table 4- Abridged Numeracy Score- Between Group Comparisons

Educational Track	n	Mean	Standard Deviation
Pre-Med	115	3.72	1.51
Medical Student	41	4.51	.93
Medical Doctor	50	4.34	.96
Pre-Law	52	3.75	1.61
Law Student	62	4.27	1.16
Lawyer	50	4.16	1.48
Engineer	71	4.06	1.30
Other	278	3.93	1.21
Total	720	4.00	1.29

*Notes: *p<0.05, **p<0.01*

Within Group comparisons

When examining the law track, our one-way ANOVA model was insignificant $F(2, 162) = 1.99$, $p = .139$, $\eta p^2 = .024$, (see Table 5). Because the ANOVA model did not reveal significant differences, no pairwise comparisons were performed. At the extremes, the means had a relatively small range from 3.75 to 4.27.

Table 5- Abridged Numeracy Score- Within group Comparison for Law track

Educational Track	n	Mean	Standard Deviation
Pre-Law	52	3.75	1.61
Law Student	62	4.27	1.16
Lawyer	50	4.16	1.48

*Notes: *p<0.05, **p<0.01*

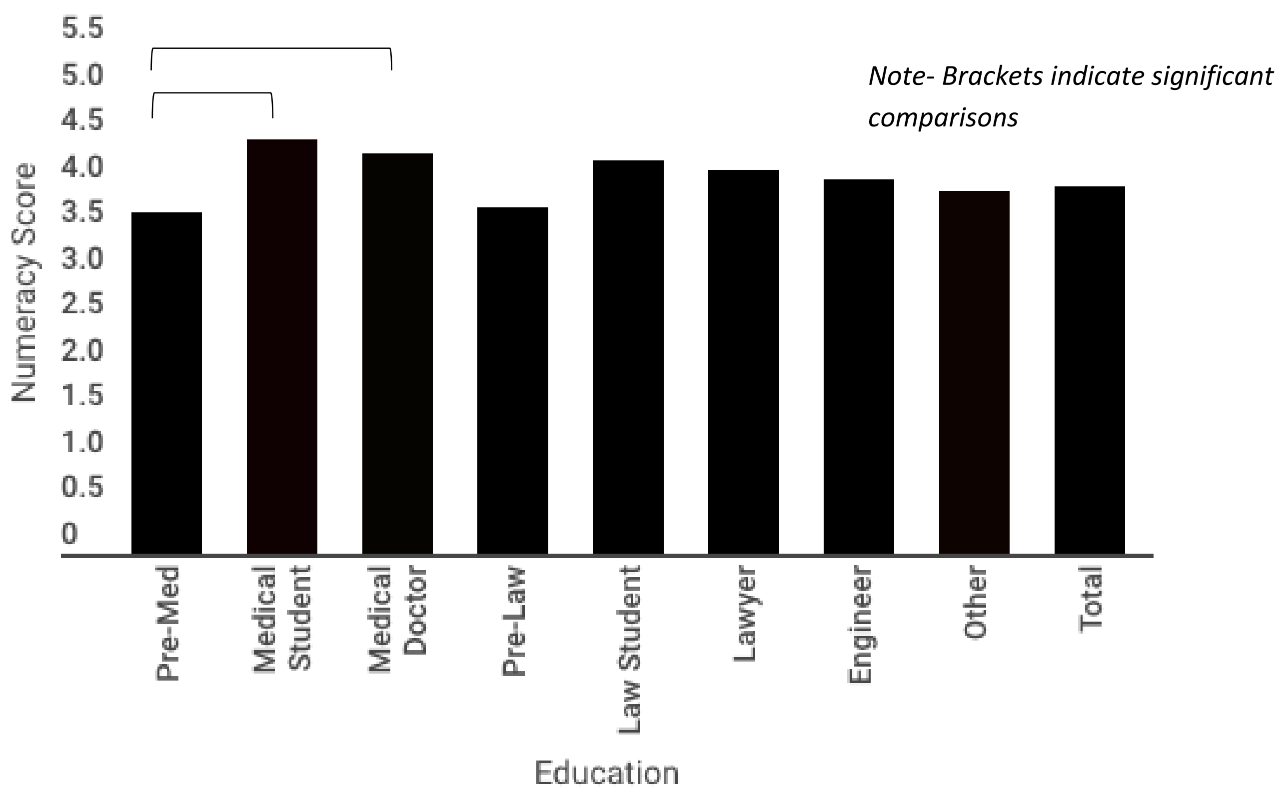
When examining the medical track on the other hand we did yield a significant difference between pre med students and their older educational counterparts using a one-way ANOVA (2, 204) = 7.23, p = .001, ηp^2 = .066 and the Bonferroni correction for pairwise comparisons. Pre meds scored lower than Medical students (p=.004) as well as medical graduates (p=0.019). There were no other comparisons that showed significant differences (see Table 6).

Table 6- Abridged Numeracy Score- Within group Comparison for Medical track

Educational Track	n	Mean	Standard Deviation
Pre-Med	115	3.72	1.51
Medical Student	41	4.51	.93
Medical Doctor	50	4.34	.96

*Notes: *p<0.05, **p<0.01*

Figure 2- Abridged Numeracy Score Comparisons



The Wason Card Task

Since the Wason Cards Task was a single question that could either be correct or incorrect, we used a binary logistic regression to determine the effect of education on performance. Specifically, Lawyers and Doctors were used as comparison groups for the analyses. First, physicians were compared, followed by lawyers to reveal between group differences, and then we examined the differences in performance within groups. When physicians were the comparison group, the omnibus test for the model was significant (p=0.002), and revealed some interesting differences. Medical students were about 3.8 times more likely to correctly answer the problem than physicians (p=0.031). Interestingly, without exception all members of the law track (pre-law, law students, and lawyers) outperformed doctors in this comparison. Pre law students were 3.4 times as likely to answer correctly (0.045), law students 5.8 times (p=0.001), and lawyers 4 times (p=0.021). Engineers also outperformed physicians 4 fold (p=0.016). No other comparisons were significant in this model (see table 7).

Table 7- Wason Cards Task- Medical Doctors as reference group

Educational Track	b	SE	Wald	Accuracy
Pre-Med	0.96	0.65	2.17	0.12
Medical Student	1.48	0.69	4.63	0.19
Medical Doctor	----	----	21.99	0.05
Pre-Law	1.37	0.69	4.00	0.17
Law Student	2.08	0.64	10.45	0.29
Lawyer	1.56	0.67	5.32	0.20
Engineer	1.58	0.62	2.30	0.12
Other	0.94	0.66	5.79	0.20

When Lawyers were used as the comparison group, the omnibus test of the model was not significant (p=0.224). There were no significant comparisons between lawyers and other professional or between lawyers and their younger educational counterparts. However, even though none of the contrasts reached significance, a trend did emerge within both the medical and law track (See Figure 2). This trend, that did reach significance in the medical track but not the law track shows the same peak performance during a student's professional education.

Within Group Comparison

When comparing performance on the Wason task within the medical profession, a binary logistic regression was used. Physicians were used as a comparison group, and the omnibus test of the model was not significant (p=0.069). Since the model was not significant, the comparisons contained within have little impact; however, it is worth noting that the same trends shown in figure 2 emerge with this model even though the comparisons were not able to reach significance (see Table 8).

Table 8- Wason Cards Task- Within Medical profession, Doctors as reference group

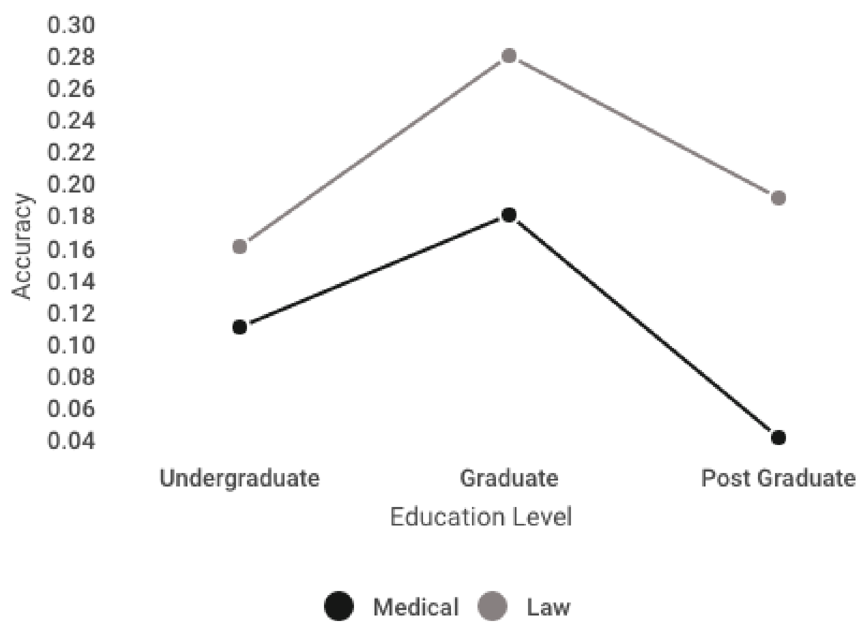
Educational Track	b	SE	Wald	Accuracy
Pre-Med	0.96	0.65	2.17	0.12
Medical Student	1.46	0.69	4.50	0.19
Medical Doctor	----	----	4.59	0.05

A binary logistic regression was used to determine the effects of legal education on Wason cards task score. Lawyers were used as the comparison group and the omnibus test of the model was not significant (p=0.224). None of the comparisons within this model reached significance, but interestingly the model does show the same trend as seen in the between group comparison- namely that law students outperform both lawyers and pre-law students (see Table 9). The latter contrast suggests that there is something inherent in legal training that improves deductive reasoning ability over what pre-law students exhibit during their undergraduate training. As to why actual attorneys did not appear to reason as well as law school students, future research with scientific sampling will be necessary.

Table 9- Wason Cards Task- Within Law profession, Lawyers as reference group

Educational Track	b	SE	Wald	Accuracy
Pre-Law	-0.09	0.46	0.04	0.17
Law Student	0.55	0.41	1.79	0.29
Lawyer	----	----	2.99	0.20

Figure 3- Wason Cards Task Within Group Comparison



The Linda Problem

Here, again, a binary logistic regression was used to compare groups on their average accuracy on the Linda problem. First, physicians were used as the comparison group. The omnibus test of the overall model was significant ($p<0.001$) and showed that lawyers were about 1.56 more likely to get the correct answer than physicians ($p=0.004$). Interestingly, physicians were 1.64 times as likely to answer correctly than those in the “other group ($p=0.007$). No other comparisons yielded significant differences (see table 10) This shows tha the two professional groups, despite sometimes differing from each other, were actually superior to members of the educated public who did not enter professional training.

Table 10- Linda Problem- Medical Doctors as reference group

Educational Track	b	SE	Wald	Accuracy
Pre-Med	-0.50	0.33	2.22	0.34
Medical Student	0.06	0.40	0.02	0.48
Medical Doctor	----	----	45.76	0.46
Pre-Law	-0.60	0.40	2.25	0.32
Law Student	0.12	0.37	0.10	0.49
Lawyer	1.22	0.42	8.39	0.75
Engineer	-0.12	0.30	0.12	0.28
Other	-0.82	0.36	7.27	0.43

When Lawyers were used as the comparison group, the binary logistic regression was significant (p<0.001). The data show that Lawyers outperformed every other group in our experiment. Lawyers were 2.21 times as accurate as Pre-med students (p<0.001), 1.56 times as accurate as medical students (p=0.008), and 1.63 times as accurate as physicians (p=0.004). Further, Lawyers were 2.34 time as accurate as pre-law students (p=0.007), and 1.53 times as accurate as Law students (p=0.007). They performed 1.74 times as accurately as people in the “other” group (p=0.001) and surprisingly 2.68 times as accurately as engineers (p<0.001) (see table 11). This suggests that there is either something specific to practicing law that sensitizes practitioners to conjunction fallacies.

Table 11- Linda Problem- Lawyers Doctors as reference group

Educational Track	b	SE	Wald	Accuracy
Pre-Med	-1.72	0.37	21.13	0.34
Medical Student	-1.16	0.44	7.06	0.48

Medical Doctor	-1.22	0.42	8.39	0.46
Pre-Law	-1.82	0.44	17.51	0.32
Law Student	-1.10	0.41	7.31	0.49
Lawyer	----	----	45.74	0.75
Engineer	-2.04	0.40	11.42	0.28
Other	-1.35	0.35	34.46	0.43

Within-Group Comparisons

A binary logistic regression was used to assess the effect of medical education on performance on the Linda Problem. Physicians were used as the comparison group and the omnibus test for the model was not significant (p=0.185). Physicians were shown to be 1.35 times as accurate as pre-medical students; however, this comparison was not significant (p=0.136). No other comparisons were statistically significant (see Table 12).

Table 12- Linda problem- Within Medical profession, Doctors as reference group

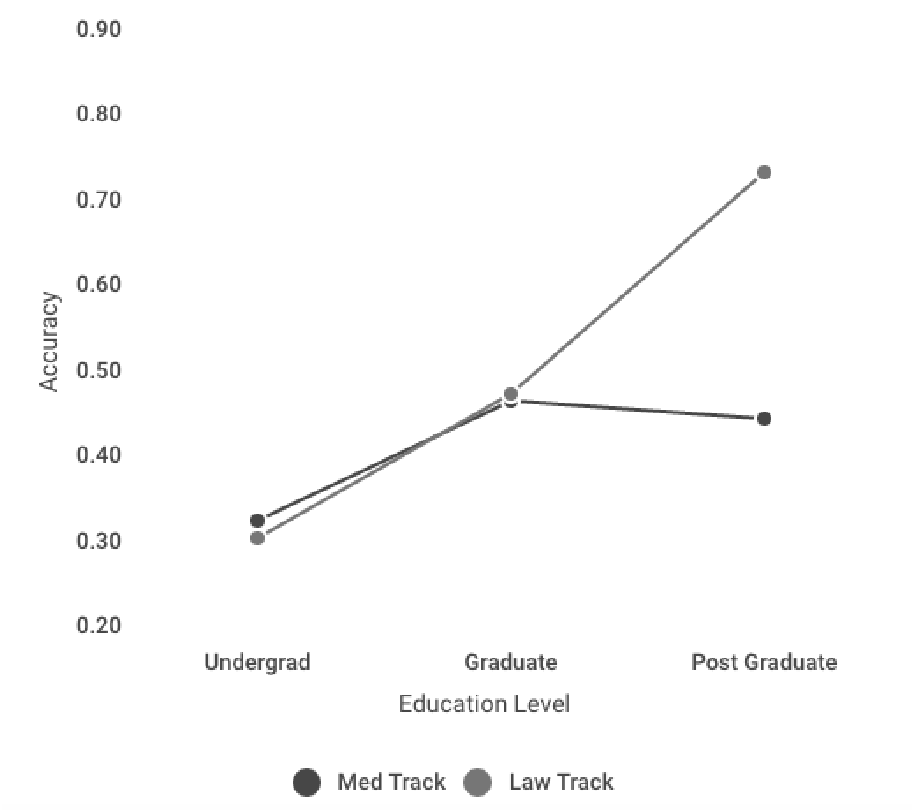
Educational Track	b	SE	Wald	Accuracy
Pre-Med	-0.50	0.33	2.22	0.34
Medical Student	0.02	0.40	0.003	0.48
Medical Doctor	----	----	3.36	0.46

A binary logistic regression was also used to compare performance on the Linda problem within the legal profession. Lawyers were used as the comparison group and the omnibus test for the model was significant (p<0.001). This model showed that Lawyers were 2.34 times as accurate than pre-law students (p<0.001) and 1.53 times as accurate as law students (p=0.01). Because accuracy increases as a function of training in the field, it suggests that law education has an impact on its students' cognitive abilities (see figure 3). Namely it shields them from the conjunction fallacy being tested here. No, other comparisons attained significance (see Table 13).

Table 13- Linda Problem- Within Law profession, Lawyers as reference group

Educational Track	b	SE	Wald	Accuracy
Pre-Law	-1.803	0.43	17.77	0.32
Law Student	-1.03	0.40	6.58	0.49
Lawyer	----	----	17.81	0.75

Figure 4- Linda Problem- Trends in within-group comparisons



Beads Task

As noted earlier, we used two different versions of the beads task. One was labeled “easy” and one was labeled hard”. The difference between the two versions was the ratio of blue to red beads in the jar. The ratio in the difficult version was closer to 1:1, thus making the decision about which jar is being chosen a more difficult task.

Easy Version

We used a one-way ANOVA test to test for effect of profession on the number of beads drawn before coming to a conclusion. The model was significant, indicating that there is a reliable effect of education on number of beads drawn $F(7, 629) = 2.11$, $p = .04$, $\eta p^2 = .023$. Unfortunately, even though the overall model was significant, none of the individual comparisons between professions reached significance (see table 14).

Table 14- Number of beads draw-, easy version- Between Group Comparisons

Educational Track	n	Mean	Standard Deviation
Pre-Med	115	3.91	3.33
Medical Student	28	4.04	3.68
Medical Doctor	48	3.83	2.56
Pre-Law	49	3.57	3.23
Law Student	31	5.03	5.87
Lawyer	42	6.00	6.15
Engineer	62	5.18	4.29
Other	262	4.43	3.90
Total	637	4.41	4.03

To test the effect of medical education specifically on the number of beads drawn, we performed another one-way ANOVA test. We sought to reveal any differences in the amount of information that different people need to make a decision at different points in their career. The ANOVA was not significant $F(2, 122) = .052$, $p = .95$, $\eta p^2 = .001$ (see Table 15).

Table 15- Number of Beads drawn-Easy Version- Within group Comparison for Med track

Educational Track	n	Mean	Standard Deviation
Pre-Med	115	3.91	3.33
Med Student	28	4.04	3.68
Doctor	48	3.83	2.56

To test the effect of legal education specifically on the number of beads drawn, we performed another one-way ANOVA test. The analysis also did not yield any significant results $F(2, 122) =2.57$, $p = .08$, $\eta p^2 = .04$, (see Table 16).

Table 16- Number of Beads Drawn-Easy Version- Within group Comparison for Law track

Educational Track	n	Mean	Standard Deviation
Pre-Law	49	3.57	3.23
Law Student	31	5.03	5.87
Lawyer	42	6.00	6.15

Difficult Version

We used a one-way ANOVA test to test for effect of profession on the amount of information gathered before coming to a conclusion, measured by the number of beads drawn. The model was significant, indicating that there is a significant effect of education on number of beads drawn $F(7, 696) = 2.89$, $p = .005$, $\eta p^2 = .028$. The overall model was significant and revealed some interesting differences between groups once the Bonferroni correction was used for pairwise comparisons. Law students drew significantly more beads than pre-med students (p=0.011) or those classified in the "other" category (p=0.046) (see table 17).

Table 17- Number of beads draw-, difficult version- Between Group Comparisons

Educational Track	n	Mean	Standard Deviation
Pre-Med	113	6.67	5.57
Medical Student	40	8.13	5.20
Medical Doctor	50	8.48	7.45
Pre-Law	50	7.22	6.06
Law Student	59	8.37	6.95
Lawyer	47	10.43	8.58
Engineer	70	9.39	5.48
Other	275	7.39	5.79
Total	704	7.87	6.14

To test the effect of medical education on the number of beads drawn in the difficult version of the beads task, we performed a one-way ANOVA test. We sought to reveal any differences in the amount of information that different people need to make a decision at different points in their medical career. The ANOVA was not significant $F(2, 202) = 2.00$, $p = .138$, $\eta p^2 = .019$ (see Table 20). (see Table 18).

Table 18- Number of Beads drawn-difficult Version- Within group Comparison for Med track

Educational Track	n	Mean	Standard Deviation
Pre-Med	113	6.67	5.57
Med Student	40	8.13	5.20
Doctor	50	8.48	7.45

To test the effect of legal education specifically on the number of beads drawn in the difficult version of this task, we performed another one-way ANOVA test. The analysis did not yield any significant results F (2, 154) = 2. 15, p = .12, ηp^2 = .027, (see Table 19).

Table 19- Number of Beads Drawn-Easy Version- Within group Comparison for Law track

Educational Track	N	Mean	Standard Deviation
Pre-Law	50	7.22	6.06
Law Student	59	8.37	6.95
Lawyer	47	10.43	8.58

Figure 5- Easy Beads Task- within-group comparisons

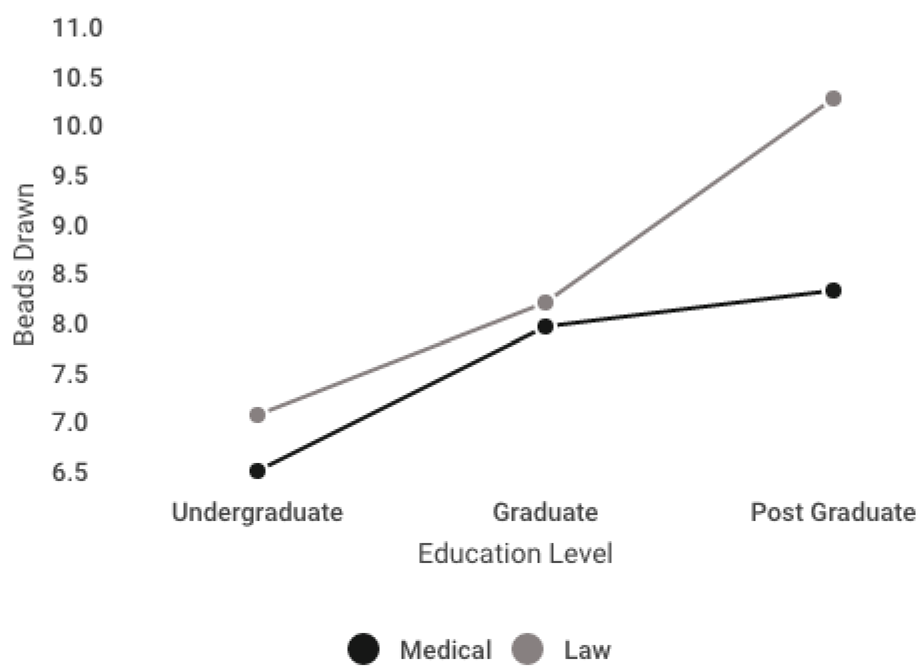
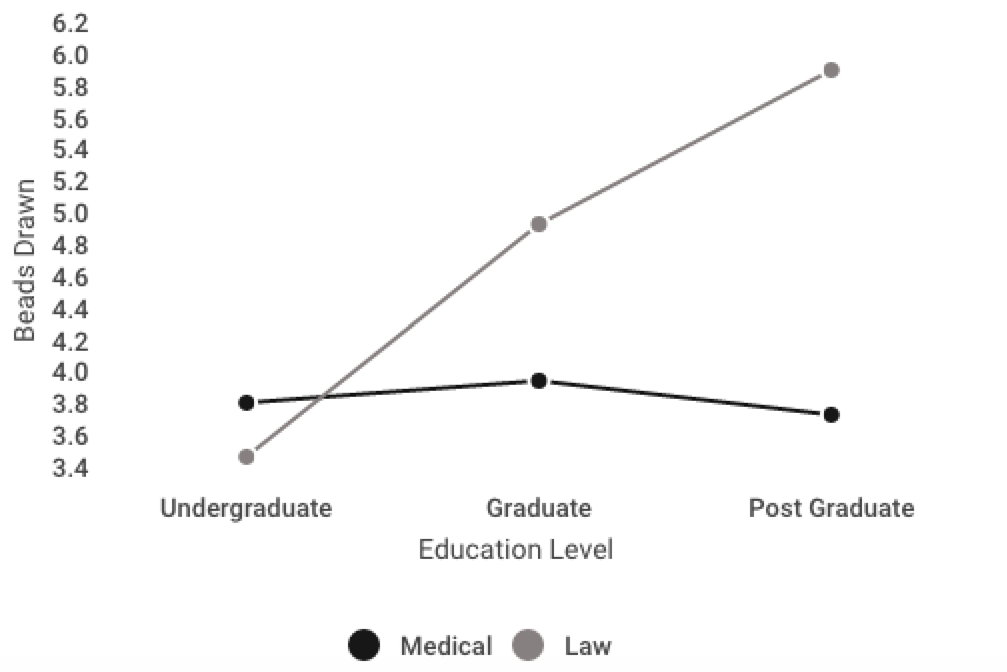


Figure 6- Difficult Beads Task- within-group comparisons



Discussion

The current study aimed to explore the effect of education on performance on a battery of cognitive tests. These tests seek to evaluate both critical thinking ability, such as deduction, as well as decision-making preferences such as how much evidence is required before leaping to a conclusion. We looked both between professions and within the legal and medical professions for differences in performance. Analyses revealed some interesting trends. First, and perhaps most surprisingly, despite their higher numbers of years of education, trained professionals (lawyers, and doctors) did not consistently outperform the general population on these critical thinking measures. There were certain instances in which lawyers performed better than the general population, however this was not as ubiquitous as hypothesized and interestingly doctors did not outperform the general population on any of the measures included. Results also indicated that our hypothesis that performance on these cognitive measures would peak in during graduate education was only partially supported. In the Cards task, the numeracy scale, and the CRT, this assumption held true, but it was not as clear in the Linda problem. More specifically this difference seems to differ between the law and medical education. In the law track, it seems that for our measures, time spent in the profession generally leads to increased performance, while for the medical track performance peaks in graduate school and declines during professional practice. Additionally, the results did not show a clear divergence in decision-making patterns between Law school graduates and Medical doctors. The differences seen between doctors and lawyers consistently show lawyers outperforming doctors, and while this difference did not always reach significance, that surprising trend did frequently emerge in the results. When exploring the tendency of different professions to jump to conclusions, no trends emerged and no comparisons yielded a significant result, which was in stark contrast to our hypothesis that the exigencies facing doctors encourage them to rely on less information in their practice.

Between-group differences in Cognitive reflection

The CRT is made to measure one's ability to halt their initial reaction, reflect, and change their answer to the less appealing, but ultimately correct answer. We predicted that doctors would have a lower CRT score than lawyers because of the necessity to make quick judgments in the course of their practice. We found, however, that medical professionals did not perform significantly worse than lawyers on that measure. They were however just as accurate as the general population. Interestingly, there were some significant differences that emerged in this comparison that we did not expect. For instance, pre-med students scored significantly lower than engineers on the CRT. This was an interesting finding, but one that we did not seek to explore. The results of our CRT analysis were surprising because of the relatively small differences between groups of people who were so differently educated. Especially considering the findings of Fredrick et at. (2005), which suggested that individuals from high-ranking universities (e.g., Harvard, MIT) were able to greatly outperform members of the general population as well as members of less elite universities (e.g., University of Toledo).

Within-group differences in Cognitive reflection

We sought to ascertain some information about the content of medical and legal education. Specifically, how exactly does education affect the way professionals employ system 1 versus system 2 decision making? On the CRT we found that pre medical students scored significantly lower than medical students, but surprisingly they seemed to perform just as well as doctors. This suggests that cognitive reflection peaks during medical school. This might indicate that this particular skill is important during a medical education but the need for cognitive reflection declines as these individuals graduate and begin to practice medicine. Another possible explanation for this surprising result is a sampling bias. It is possible that the doctors surveyed were not as adept as the pre-medical students we surveyed, perhaps because of changes to the medical curricula over time. Lawyers on the other hand seem to be able to retain improvements on the CRT made during schooling well into their career. Pre-law students scored lower than law students, but lawyers were able to score just as well as law students indicating that there was not a decline post graduation in cognitive reflection. Further research in this area should focus on a larger and more diverse cohort of individuals on both the legal and medical track, not only to rule out any sampling biases but also to improve the statistical power of the comparisons made here.

Between-group differences in deductive and statistical reasoning

Using the Wason Cards task and the Linda problem to test deductive reasoning and statistical reasoning respectively, we compared different professions in our cohort. Here doctors scored higher than the general population on the Linda problem, but were unable to score higher than the general population the deductive reasoning task. Lawyers were able to score higher than all other groups studied on the Linda problem indicating a high level of dexterity with statistical reasoning. Lehman and Nisbett (1990) show that professions in social science and psychology are generally more adept with statistical reasoning than individuals in the natural sciences. It is not unreasonable to assume that lawyers would more likely have a background in social science or psychology than doctors and this could partially explain the result shown. Lawyers were only able to outperform doctors on the deductive reasoning task, but this might simply be because of their frankly dismal (0.05 accuracy rate) on that particular measure.

Within-group difference in deductive and statistical reasoning

Looking within each profession studied we found that doctors and pre-med students performed comparably on the Wason cards task while doctors were more accurate on the Linda problem than Pre-med students. Current medical school students were superior to medical school graduates on the deductive reasoning task and performed comparably on the Linda problem. Barring sampling bias, these results lend some more partial support to the idea that certain critical thinking abilities decline in medical professionals after graduation. Lawyers were more accurate than law students, who were more accurate than pre-law students on the Linda problem. The linear trend shown in figure 3 depicts a steady increase in statistical reasoning ability throughout legal education. On the deductive reasoning task, there was a trend that is more characteristic of our hypothesized peak in thinking ability during graduate education. Although none of the comparisons within the legal track reached significance, the trend that emerges is one where current law students outperformed both pre-law students and lawyers. Once again, the issue of sampling becomes relevant and future research by someone with greater resources than I will be needed to insure scientific sampling.

Between-group differences in evidence based decision making

We hypothesized that doctors would require less information than lawyers based on the need for quick decision making in the practice of medicine. This hypothesis however was not substantiated by the data.

In fact, on this dimension we found no significant differences between any of the groups studied. This however was due to a very high variability within the data set. It is worth noting however, that while these differences did not reach significance, lawyers on average were more inclined to draw more beads on both the easy and hard version of the task. Perhaps this is the result of the risk-aversive nature of many legal decisions; unlike medicine in which a practitioner does not have the luxury of waiting until all evidence is collected before initiating treatment, lawyers are aware of the consequences of decisions that are irreversible, thus delaying making them until all options are considered.

Within-group differences in evidence based decision making

Our exploratory analysis revealed some interesting trends within each professional track studied. While even these within group trends did not reach significance, they were nonetheless informative and something to consider when looking toward future research. First, we found that within the medical track, there was no obvious difference in preference for amount of information gathered in the beads task. Within the law profession however, it seemed that as legal education continued, a desire for more concrete evidence emerged (see figure x).

Between-group numeracy comparisons

When comparing the legal and medical profession in their ability to understand numerical problems, we found that only pre-medical students were significantly worse than medical students. No other comparisons yielded significant differences, however a trend emerged in which both medical and legal graduates had a slight edge over the general population. Interestingly, when we subdivided the general population to examine only those who were engineers we found that they performed no better than any other group. This is a surprising result given the extensive mathematical background of most engineers.

Within-group numeracy comparisons

We hypothesized that graduate education would improve dexterity with numbers and thus improve the score on our numeracy scale. This hypothesis was not entirely borne out in the data however because we only found that medical students and doctors outperformed pre-med students. This trend was not seen with lawyers, indicating that their training may not emphasize math or the ability to handle numbers well.

Conclusion

The results from this study focused on both inter- and intra- professional differences. There have been many studies published that have sought to quantify the differences between groups of individuals on various measures including the ones used here. What makes this study unique is its ability to look at a cross-section of the educational track of both lawyers and doctors. We sought to shed light on a question of self-selection. We wanted to explore whether certain cognitive profiles were present in individuals prior to the onset of their education or if the very nature of their education is what shaped their cognitive profiles. Our results don't necessarily give us a clear answer to this question. For instance, on one hand with the Linda problem (a measure of statistical reasoning) there seems to be evidence to suggest that legal education significantly improves performance in this area of critical thinking (see figure 2). On the other hand, however, Lawyers do not seem to keep the cognitive gains made in cognitive reflection following their law school years. In the same vein, physicians seem to lose any cognitive gains made during their medical schooling on basically all of the measures used here, and on the measures where they manage to avoid this decline, it is simply because they perform identically to pre-med and medical students, not outperform them. As explained below, I do not believe, however, that these results are contradictory.

Given that critical thinking is a broad and very nuanced aspect of cognition, it is entirely possible that the current study did not manage to paint an accurate picture of the skills taught during legal and medical education. The fact that the cognitive properties tested here are very relevant to both fields, it does not mean that they are necessarily emphasized in either profession's curriculum. This would then help explain the incomplete educational phenomena we have begun to describe in the current study. Another possible explanation for these sporadic results are that medical professionals, medical students, and pre-medical students come from a less educationally diverse pool of individuals. Since the requirements to attend medical school are strict and rigorous, only a certain type of person is willing to embark on that path. This would then help explain why medical school does not seem to alter the cognitive profile of its students: the individuals that were not suited to that style of thinking were either not attracted to the profession or eliminated from the potential pool early in their baccalaureate years. With regard to law school, however, the requirements are decidedly less defined. Students from any major can be "pre-law" as there is no required coursework that is pre-requisite for attending law school. Thus, the pool of students who identify as "pre-law" and even those who go on to law schooling will by definition be from a group of students who are more diverse thinkers, or at the very least are from educational backgrounds that foster different styles of thinking. When these students are churned through the mill of legal education, their patterns of thinking inevitably become more uniform as now they have all undergone the same training. This could explain why law school and practicing law seem to have a larger impact on critical thinking measures that medical schooling, and it is consistent with the generally higher SDs observed among members of the legal community (see Beads Task SDs in Tables 17-19). Notwithstanding this, the trend of a decline in cognitive skills among doctors and the general underperformance of the individual's medical track is of note and concern.

Future research should seek to repeat this study with the aim of identifying more nuanced differences in performance between lawyers and doctors. First, this study would need a much larger sample size than the one used in the current study, and scientific sampling employed—a luxury not possible in the present study due to limited resources. Additionally, snowball sampling was used to obtain professional and graduate participants which meant that each group we compared was relatively homogeneous with regard to educational background, SES, and ethic background. This homogeneity might help explain some of the non-significant differences we found between groups. Additionally, in order to make the results here more applicable to a broader population, future research should seek to examine each year of schooling separately as well as examining the difference between different specialties. This way we would be able to get a better understanding of when cognitive skills improve and decline and under what specific educational circumstances this happens. This way, professors and curriculum designers may be able to shape the education they bestow on their students to better prepare them with critical thinking skills. It may also be useful to compare the education attained at different universities because it is known that quality of education has an impact performance on many of the measures used above (e.g., Frederick, 2005).

Unfortunately, although the measures used in the current study were previously validated and have high internal consistency, they do not test all aspects of critical thinking. Due to constraints on participant attention we were unable to include as many measures as we would have liked and thus future research would certainly benefit from including a wider array of cognitive measures in order to get a more complete cognitive profile for each participant. Additionally, some of the measures used in this study did not perform as well as we would have hoped. The numeracy scale particularly was problematic in this study because of the extremely high accuracy rate among all groups, possibly obscuring subtler differences that went undetected. The beads task was similarly crude due to a very high variability within each group (often the standard deviation was as high as the mean number of beads chosen). Thus, this measure had relatively little explanatory value. Future studies should certainly consider using more sensitive and reliable tests to tease out any effects that we may have missed due to the rough nature of the tools used here.

This study sought to examine how individuals in certain careers were shaped by their education. There is an important distinction between selecting a career based on existing cognitive traits or having that career mold a certain cognitive profile. This distinction also has important implications for the thousands of people who sit at a crossroads every year when they decide what career they will pursue for the rest of their lives. Additionally, it is important to understand how our education shapes the way we think so as to more responsibly wield that powerful tool to give students the best cognitive advantages that will help them in their career. We find that the answer to this question, as we probably could have expected, is nuanced and not entirely clear. What we do know though is that most likely, both innate cognitive traits and education combine to shape the way we think. Either way, it is imperative that we delve deeper into understanding how formal education may affect cognitive reasoning.

References

Hofmans, J., & Mullet, E. (2013). Towards unveiling individual differences in different stages of information processing: a clustering-based approach. *Quality and Quantity*, *47*(1), 455–464. https://doi.org/http://dx.doi.org/10.1007/s11135-011-9529-7

Wason, P.C. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology, 20,* 273 – 281. doi: 10.1080/14640746808400161

Warman, D.M., Lysaker, P.H., Martin, J.M., Davis, L., & Haudenschield, S.L. (2007). Jumping to conclusions and the continuum of delusional beliefs. *Behaviour Research and Therapy*, *45*, 1255– 1269. doi: 10.1016/j. brat.2006.09.002

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy and probability judgment. *Psychological Review, 90,* 293 – 315. doi:http://dx.doi.org/10.1037/0033-295X. 90.4.293

Neth, H., & Gigerenzer, G. (2015). Heuristics: Tools for an uncertain world. In R. Scott & S.

Kosslyn (Eds.), Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource (pp. 1 – 18). New York, NY: Wiley Online Library. doi: 10.1002/9781118900772.etrds0394

Nakao, K., & Treas, J. (1992). The 1989 socioeconomic index of occupations: Construction from the 1989 occupational prestige scores. Chicago: National Opinions Research Center.

Mandal, A., Ghosh, A., Sengupta, G., Bera, T., Das, N., & Mukherjee, S. Factors affecting the performance of undergraduate medical students: A perspective. Indian Journal of Community Medicine: Official Publication of Indian Association of Preventive & Social Medicine, 37, 126-129. doi: 10.4103/0970-0218.96104

Loftus, E.F., Levidow, B., & Duensing, S. (1992). Who remembers best? Individual differences in memory for events that occurred in a science museum. Applied Cognitive Psychology, 6, 93 – 107. doi: 10.1002/acp.2350060202

Lipkus, I.M., Samsa, G., & Rimer, B.K. (2001). General performance on a numeracy sale among highly educated samples. *Medical Decision Making, 21,* 37 – 44. doi: 10.1177/0272989X0102100105

Liberali, J.M., Reyna, V.F., Furlan, S., Stein, L.M., & Pardo, S.T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making, 25,* 361 – 381. doi: 10.1002/bdm.752.

Huq, S.F., Garety, P.A., & Hemsley, D.R. (1988). Probabilistic judgments in deluded and non- deluded subjects. The Quarterly Journal of Experimental Psychology, 40, 801 – 812. doi: 10.1080/14640748808402300

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. Academic Medicine, 73, 538 – 540. doi: http://dx.doi.org/10.3389/fpsyg.2015.00642

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62,* 451 – 482. doi: 10.1146/annurev-psych-120709-145346

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19,* 25 – 42. doi: 10.1257/089533005775196732

Evans, S., Almahdi, B., Sultan, P., Sohanpal, I., Brandner, B., Collier, T., Shergill, S.S., Cregg, R., & Averbeck, B.B. (2012). Performance on a probabilistic inference task in healthy subjects receiving ketamine compared with patients with schizophrenia. Journal of Psychopharmacology, 26, 1211 – 1217. doi: 10.1177/0269881111435252

Elstein, A. S., & Schwarz, A. (2002). Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *British Medical Journal*, *324*, 729 – 732.

Jacoby, R. J., Abramowitz, J. S., Buck, B. E., & Fabricant, L. E. (2014). How is the Beads Task related to intolerance of uncertainty in anxiety disorders? Journal of Anxiety Disorders, 28(6), 495–503. https://doi.org/10.1016/j.janxdis.2014.05.005

Evans, J. S. B. T., & Ball, L. J. (2010). Do people reason on the Wason selection task? A new look at the data of Ball et al. (2003). Quarterly Journal of Experimental Psychology (2006), 63(3), 434–441. https://doi.org/10.1080/17470210903398147

Berwick, D. M., Fineberg, H. V., & Weinstein, M. C. (1981). When doctors meet numbers. The American Journal of Medicine, 71(6), 991–998.

Franken, I. H. A., & Muris, P. (2005). Individual differences in decision-making. Personality and Individual Differences, 39(5), 991–998. https://doi.org/10.1016/j.paid.2005.04.004

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach. Journal of Behavioral Decision Making, 26(2), 198–212. https://doi.org/10.1002/bdm.1751