

BETSY COOPER

## Judges in *Jeopardy!*: Could IBM's Watson Beat Courts at Their Own Game?

### INTRODUCTION

February 16, 2011 was a day of reckoning for humankind. A new computer, appropriately dubbed “Watson,” beat the world’s best *Jeopardy!* players at their own game. At first blush this may not seem so surprising: after all, computers are notoriously better than humans at “recalling” factual knowledge. But *Jeopardy!* is a game show known for the nuance of its clues, which often contain puns, ambiguities, and other curiosities.<sup>1</sup> Watson’s ability to understand and quickly respond to *Jeopardy!* questions thus reveals that computers have made great strides in emulating how humans think.

Watson is a computer built for a very specific purpose: to beat humans at *Jeopardy!*. Since his victory, pundits and IBM staffers have suggested that the technology powering Watson might have many uses—in the gaming world, for example, or improving customer service from much-maligned automated call centers.<sup>2</sup> Only a week after winning the *Jeopardy!*

- 
1. The majority of *Jeopardy!* clues are based on factual information. Some clues “contain multiple facts about the answer, all of which are required to arrive at the correct response but are unlikely to occur together in one place.” Still others include word puzzles; in the “Rhyme Time” category, the respondent must present two sub-clue answers that rhyme with each other. David Ferrucci et al., *Building Watson: An Overview of the DeepQA Project*, AI MAG., Fall 2010, at 59, 60, available at <http://www.stanford.edu/class/cs124/AIMagazine-DeepQA.pdf>.
  2. See, e.g., Stephen Baker, *Watson Is Far from Elementary*, WALL ST. J., Mar. 14, 2011, <http://online.wsj.com/article/SB10001424052748704570904576180954262138300.html>; Dalibor Dimovski, *Does Watson Herald the Future of Multiplayer Gaming?*, SIDEQUESTING, <http://www.sidequesting.com/2011/02/does-watson-herald-the-future-of-multiplayer-gaming> (last visited Aug. 23, 2011); Chad Lyne, *It's Elementary, My Dear*

title, Watson's creators proclaimed to the annual Healthcare Information and Management Systems Society meeting that "Watson could dramatically improve health care delivery by offering, in minimal time, solutions that have a high level of certainty."<sup>3</sup> Here I propose how Watson could apply his skills in a legal environment<sup>4</sup>: by helping textualists interpret statutes. New textualists believe in reducing the discretion of judges in analyzing statutes. Thus, they advocate for relatively formulaic and systematic interpretative rules. How better to limit the risk of normative judgments creeping into statutory interpretation than by allowing a computer to do the work?

This Essay considers whether judges might share the job of statutory interpretation with computers like Watson. First, it briefly lays out how new textualists approach statutory interpretation. Second, it describes how Watson's aptitudes lend themselves to textualist-style statutory interpretation. Finally, the Essay pulls the threads together, discussing how Watson might both aid textualist interpretation and perhaps perform such interpretation on his own.

## I. THE NEW TEXTUALIST ASPIRATION

New textualism is a popular method of interpretation by which judges decipher statutes; perhaps its foremost proponent is U.S. Supreme Court Justice Antonin Scalia. New textualist premises and lines of reasoning about statutory interpretation have become widespread. Indeed, there is arguably a growing consensus that we (or at least judges) "are all textualists now."<sup>5</sup> Many others have outlined the key tenets of this method, so I will not spend much time doing so here.<sup>6</sup> For the purposes of this Essay, there are three important elements of new textualism: its reliance on ordinary

---

*Watson*, INTERACTIVE INTELLIGENCE (Apr. 5, 2011), <http://callcenterinfo.tmcnet.com/analysis/articles/161434-its-elementary-my-dear-watson.htm>.

3. Pamela Lewis Dolan, *Fresh from "Jeopardy!" Victory, Watson To Take on Health Care*, AM. MED. NEWS, Mar. 14, 2011, <http://www.ama-assn.org/amednews/2011/03/14/bisa0314.htm>.
4. After this article was drafted, a newsletter suggested that Watson might someday play a different type of legal role—serving as an expert witness. See Robert Ambrogi, *Could IBM's Watson Make Experts Obsolete*, BULLSEYE, April 2011, <http://www.ims-expertservices.com/newsletters/april/could-watson-make-experts-obsolete-041211.asp>.
5. Jonathan R. Siegel, *Textualism and Contextualism in Administrative Law*, 78 B.U. L. REV. 1023, 1057 (1998); see also Jonathan T. Molot, *The Rise and Fall of Textualism*, 106 COLUM. L. REV. 1, 29 (2006) (noting that "textualism has met with considerable success over the last two decades").
6. See, e.g., William N. Eskridge, Jr., *The New Textualism*, 37 UCLA L. REV. 621 (1990).

meaning (the premise), its emphasis on context (the process), and its rejection of normative biases (the reasoning). I consider each in turn.

First, new textualism begins from the *premise* that “the apparent plain meaning of a statutory text must be the alpha and the omega of a judge’s interpretation of a statute.”<sup>7</sup> The goal of textualist statutory interpretation “is to identify the objective meaning of statutory text without regard to what any legislator intended that text to mean.”<sup>8</sup> This stands in stark contrast to intentionalists, who believe that the goal of statutory interpretation should be for “courts to implement the intent of the legislature.”<sup>9</sup>

Second, the new textualist *process* of analyzing statutes takes into account the context in which a word presents itself, including the structure and coherence of the statute.<sup>10</sup> New textualists thus distinguish themselves from strict constructionists, who refuse to look at any sources outside the text of the statute.<sup>11</sup> As Justice Scalia has suggested, “when you ask someone, ‘Do you use a cane?’ you are not inquiring whether he has hung his grandfather’s antique cane as a decoration in the hallway.”<sup>12</sup> One can only understand that the question is asking whether the individual uses a cane for walking by considering it in context.

Finally, new textualists’ *reasoning* for undertaking this scheme of interpretation is to reduce the discretion that judges use when interpreting statutes. Justice Scalia warns that “the main danger in judicial interpretation . . . is that the judges will mistake their own predilections for the law.”<sup>13</sup> To avoid such errors, new textualists believe that “the goal of statutory interpretation is to determine the objective meaning of statutory

- 
7. WILLIAM N. ESKRIDGE, JR., PHILIP P. FRICKEY & ELIZABETH GARRETT, *CASES AND MATERIALS ON LEGISLATION* 779 (4th ed. 2007) (summarizing the views put forward in ANTONIN SCALIA, *A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW* 106-07 (Amy Guttman ed., 1997)).
  8. Jonathan R. Siegel, *The Inexorable Radicalization of Textualism*, 158 U. PA. L. REV. 117, 118 (2009).
  9. *Id.* at 119.
  10. See ANTONIN SCALIA, *A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW* 37 (Amy Guttman ed., 1997) (“In textual interpretation, context is everything . . .”).
  11. See *id.* at 23 (“Textualism should not be confused with so-called strict constructionism, a degraded form of textualism that brings the whole philosophy into disrepute.”).
  12. *Id.* at 24.
  13. Antonin Scalia, *Originalism: The Lesser Evil*, 57 U. CIN. L. REV. 849, 863 (1989); see also Paul Killebrew, Note, *Where Are All the Left-Wing Textualists?*, 82 N.Y.U. L. REV. 1895, 1898 (2007) (“Textualists conceive of their methodology as a bulwark against the pernicious influence of a judge’s personal politics.”).

text.”<sup>14</sup> New textualism aspires to such objectivity by advocating a relatively mechanical process of textual interpretation, divorced from the intent of Congress. Whereas traditional textualists allowed “strongly contradictory legislative history” to trump the plain meaning of a statute,<sup>15</sup> new textualists believe that “statutes [should] be read with a strict literalism and with reference to well-established canons of statutory construction” because doing so will encourage Congress to draft laws more clearly in the first place.<sup>16</sup> New textualists accordingly reject the use of legislative history as a means of understanding statutes.<sup>17</sup> Such interpretative parsimony is a perceived strength of the new textualists’ method.<sup>18</sup>

A range of critiques has been levied at new textualism. Not least, it has been accused of seeking to achieve impossible goals. After all, human judges will always begin from their own inherently subjective frame of reference.<sup>19</sup> Further, one might quibble about whether the actual practice of new textualism achieves its stated goal of eliminating bias. Some purists have argued that to take seriously the goal of eliminating bias would require rejecting particular canons of statutory interpretation, such as absurd results or scriveners’ errors.<sup>20</sup> This is in part because, by providing an escape hatch from strict textual meaning, such canons improperly allow judicial discretion to creep in.

My point here is not to dispute the importance or viability of new textualism as a mechanism for statutory interpretation. Rather, taking new textualism as a starting point—a goal, if you will—for understanding statutes, I query here whether humans, with all our cognitive biases and normative bents, are the actors best equipped to interpret statutes in this

---

14. Siegel, *supra* note 8, at 123.

15. Eskridge, *supra* note 6, at 624.

16. *Id.* at 677.

17. See SCALIA, *supra* note 10, at 29-37.

18. Professor Eskridge describes Adrian Vermeule’s “newer and more parsimonious textualism,” which advocates for greater deference to agency interpretations, as “no frills textualism.” William N. Eskridge, Jr., *No Frills Textualism*, 119 HARV. L. REV. 2041, 2043 (2006) (reviewing ADRIAN VERMEULE, *JUDGING UNDER UNCERTAINTY: AN INSTITUTIONAL THEORY OF LEGAL INTERPRETATION* (2006)).

19. Cf. WILLIAM N. ESKRIDGE, JR., *DYNAMIC STATUTORY INTERPRETATION* 204 (1994) (arguing that “dynamic statutory interpretation” requires judges “to test their own preconceptions . . . against the operation of their interpretation of the world”).

20. John Manning has suggested that rejecting these canons might display greater fidelity to textualism, though key proponents of new textualism—such as Justice Scalia and Judge Easterbrook—continue to apply them. See John F. Manning, *The Absurdity Doctrine*, 116 HARV. L. REV. 2387, 2393, 2419 n.122 (2003).

manner. It has long been recognized that humans draw poor causal inferences, especially when making judgments under uncertain conditions.<sup>21</sup> And all humans have normative biases that can confound any effort to apply interpretative rules strictly or narrowly. This leads us to the new textualist dilemma: can humans ever really be successful textualists? To answer this question, we look to Watson for a little help.

## II. “IT’S ELEMENTARY, WATSON”: *JEOPARDY!* AND STATUTORY INTERPRETATION

Watson was designed with a single goal in mind: to beat humans at their own game, *Jeopardy!*. To determine whether Watson can successfully interpret statutes, one first must understand how he functions as a *Jeopardy!* contestant. This Part considers the obstacles Watson overcame to become *Jeopardy!* champion. It also investigates how Watson answers questions, in order to see whether his methods might help resolve the new textualist dilemma.

### A. “*Who Is . . . Watson?*”

Watson is a computer system designed—as a first objective—to answer trivia questions on the game show *Jeopardy!*. In creating Watson, IBM began from the premise that computers are not, as a general principle, particularly good at answering even direct queries:

Search engines don’t answer a question—they deliver thousands of search results that match keywords. University researchers and company engineers have long worked on question answering software, but the very best could only comprehend and answer simple, straightforward questions (How many Oscars did Elizabeth Taylor win?) and would typically still get them wrong nearly one third of the time.<sup>22</sup>

Worse still, to win at *Jeopardy!*, Watson needed to be able to answer questions where important search terms were not provided. For example, *Jeopardy!*’s “decomposition”-type questions require the contestant to

---

21. See generally Amos Tversky & Daniel Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, 185 *SCI.* 1124, 1124 (1974) (showing how the use of heuristics in decisionmaking can “lead to severe and systematic errors”).

22. *A Computer Called Watson*, IBM.COM, <http://www.ibm.com/ibm100/us/en/icons/watson> (last visited Aug. 23, 2011).

“decompose the question into . . . two parts and [identify] answers to each one.”<sup>23</sup> Often, “the answer common to both questions is the answer to the original clue.”<sup>24</sup> Existing search engines like Google—which require the user to input search terms and then make use of algorithms to find instances where those terms relate most closely to one another—were ineffective at answering such questions.<sup>25</sup>

This task is made even more difficult when we consider the construction of an average “question” on *Jeopardy!* First, unlike most trivia games, *Jeopardy!* provides an *answer* and requires players to respond with the corresponding *question*—a complicated twist for a computer. And second, *Jeopardy!* questions are replete with puns, word games, and even jokes, requiring far more layers of understanding than simple recall.

To understand how this works in practice, take a sample *Jeopardy!* question: “It means detestable or loathsome, though I have no beef with the snowman, myself.”<sup>26</sup> Unlike most typical trivia questions, this *Jeopardy!* question contains no dates, facts, or even substantive knowledge. It relies instead on wordplay. You must know that a synonym for detestable is ‘abominable’ and then connect that phrase to the folklore of the snowman in the Himalayan Mountains. The average human can synthesize these streams of knowledge simultaneously, putting them together to reach the answer: “What is abominable?” But the average computer can only respond correctly if nearly the exact same question and answer appeared together in text it has learned. When I first plugged that sample *Jeopardy!* question into Google, the top hit (excluding the article from which I borrowed the question) was a blog post about wearing red high heels to a pig roast party.<sup>27</sup>

---

23. Ferrucci et al., *supra* note 1, at 62.

24. *Id.*

25. Google employs an “inverted index that maps each query word to a matching list of documents (the hit list). The index servers then determine a set of relevant documents by intersecting the hit lists of the individual query words, and they compute a relevance score for each document.” Luiz André Barroso, Jeffrey Dean & Urs Hölzle, *Web Search for a Planet: The Google Cluster Architecture*, IEEE MICRO, Mar.-Apr. 2003, at 22, 23, available at [http://static.googleusercontent.com/external\\_content/untrusted\\_dlcp/labs.google.com/en/us/papers/googlecluster-ieee.pdf](http://static.googleusercontent.com/external_content/untrusted_dlcp/labs.google.com/en/us/papers/googlecluster-ieee.pdf).

26. Adam Hadhazy, *3 Strengths and Weaknesses of the Watson Computer on “Jeopardy!”*, LIVESCIENCE.COM (Feb. 14, 2011, 5:55 PM), <http://www.livescience.com/12859-3-strengths-weaknesses-watson-computer-jeopardy.html>.

27. Jeopardy Question Query, GOOGLE.COM, <http://www.google.com> (search “It means detestable or loathsome, though I have no beef with the snowman, myself” without quotation marks) (last searched Mar. 15, 2011). The article uncovered was *Red High Heels, Slaughtered Pigs & Why Being Unsure Is a Good Thing*, THE MIDDLE

Watson's creators sought to design a computer system that could better approximate the human approach to asking and answering questions—both to determine the most likely answer and to express a level of confidence that the answer is correct. As IBM puts it, the goal of designing Watson was “to understand the actual meaning behind words, distinguish between relevant and irrelevant content, and ultimately demonstrate confidence to deliver precise final answers.”<sup>28</sup> In February 2011, IBM declared success. During a *Jeopardy!* television special, Watson took on the two most decorated human *Jeopardy!* players of all time: Ken Jennings, who once won seventy-four straight games on the show, and Brad Rutter, the player with the all-time highest prize earnings.<sup>29</sup> Watson amassed thousands of dollars in a resounding victory, which he secured even before the final round.<sup>30</sup>

After the fact, some criticized Watson's hair-trigger buzzer system, which could respond in as quickly as one-tenth of a second if he was confident in his answer. It is true that Watson beat his competitors to the buzzer in the vast majority of questions. But by any measure, Watson's victory was decisive. Watson finished the contest with \$77,147 in prize winnings. His next closest competitor, Jennings, earned only \$24,000.<sup>31</sup> Jennings responded graciously in the face of certain defeat, scrawling under his final answer in the contest that “I, for one, welcome our new computer overlords.”<sup>32</sup>

### *B. How Watson Answers Questions*

How did Watson achieve this resounding victory? IBM staff described the core components that enable Watson to answer *Jeopardy!* questions:

---

FINGER PROJECT, <http://www.themiddlefingerproject.org/you-dont-know-jack-son-red-high-heels-slaughtered-pigs-why-being-unsure-is-a-good-thing> (last visited Aug. 23, 2011).

28. *A Computer Called Watson*, *supra* note 22. The probabilistic estimation was particularly important in *Jeopardy!* because Watson would lose money if his answer was wrong. Thus, IBM designed Watson so he could decline to buzz in where he lacked confidence in his answer. *Id.*
29. Alex Strachan, *What Is a Blowout? Computer Destroys Humans at Jeopardy!*, CANADA.COM, Feb. 16, 2011, <http://www.canada.com/entertainment/What+blowout+Computer+destroys+humans+Jeopardy/4298139/story.html>.
30. See John Markoff, *Computer Wins on ‘Jeopardy!’: Trivial, It’s Not*, N.Y. TIMES, Feb. 16, 2011, <http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html>.
31. *Id.*
32. *Id.*

---

Watson runs on a cluster of Power 750™ computers—ten racks holding 90 servers, for a total of 2880 processor cores running DeepQA software and storage. It can hold the equivalent of about one million books worth of information. . . .

When a question is put to Watson, more than 100 algorithms analyze the question in different ways, and find many different plausible answers—all at the same time. Yet another set of algorithms ranks the answers and gives them a score. For each possible answer, Watson finds evidence that may support or refute that answer. So for each of hundreds of possible answers it finds hundreds of bits of evidence and then with hundreds of algorithms scores the degree to which the evidence supports the answer. The answer with the best evidence assessment will earn the most confidence. The highest-ranking answer becomes the answer. However, during a *Jeopardy!* game, if the highest-ranking possible answer isn't rated high enough to give Watson enough confidence, Watson decides not to buzz in and risk losing money if it's wrong. The Watson computer does all of this in about three seconds.<sup>33</sup>

For Watson to successfully play *Jeopardy!*, his creators relied on the *premise* that computers are better than humans at storing data. For the *Jeopardy!* challenge, “the sources for Watson include[d] a wide range of encyclopedias, dictionaries, thesauri, newswire articles, literary works, and so on.”<sup>34</sup> Watson can access a good proportion of knowledge in the public sphere, both colloquial and expert, and has perfect retention as long as he can access the information. He is thus well equipped to understand what information is available, and importantly, how frequently that information appears.

Watson *processes* this vast array of information by looking for relationships between the clue and other words; he figures out what words mean in context. Unlike previous computers, Watson can sort out the most relevant words in a clue and better target his answer. He identifies the “focus of the question,” detects relationships between words in the question, and decomposes questions into sub-questions, among other techniques.<sup>35</sup> In the question about the abominable snowman, for example, Watson might downplay the closeness of words such as “loathsome,” “detestable,” and “myself”—the apparent reason why the blog post on red

---

33. *A Computer Called Watson*, *supra* note 22.

34. Ferrucci et al., *supra* note 1, at 69.

35. *Id.* at 70.

heels and pigs was elevated by the Google algorithm. Instead, Watson might look for the links that connect “detestable” to “snowman,” which would be far more likely to produce the right result. Even better, Watson can learn from his mistakes through trial and error; he stores incorrect answers and incorporates them into future games. This helps explain how Watson went from losing in trial *Jeopardy!* competitions to beating the top-ranked competitors of all time.<sup>36</sup>

Finally, Watson is not subject to some key *reasoning* errors of humans. Of course, this is in part because Watson is subject to no normative inclinations of his own; he is only biased insofar as the human-controlled inputs fed into Watson’s memory are biased. Likewise, in contrast to humans who are notoriously poor at estimating probabilities, especially in the face of irrelevant information,<sup>37</sup> Watson is able to express the probability that he is right in a systematic and quantifiable way. More specifically, Watson can estimate how likely it is that a particular answer he provides is correct and refrain from responding to a question if the likelihood is small (by refusing to buzz in). These skills, as I argue in the following Part, get to the essence of new textualist approaches to statutory interpretation.

### III. WATSON: THE NEW TEXTUALIST?

Watson has many potential applications—and perhaps not just for search engines and scientists. Watson-style computers already help lawyers sift through documents in discovery and decipher patterns in clients’ activity.<sup>38</sup> At least one judge already accepts that judges are “not like the supercomputer Watson. . . . [T]hey have no hope of knowing everything.”<sup>39</sup> But could Watson and judges work together and revolutionize statutory interpretation? This Part considers, first, whether Watson could perform certain tasks of new textualism better than judges; and second, whether he might somehow assist (but not replace) them in performing such tasks.

---

36. *See id.* at 76 fig.9, 78.

37. *See, e.g.,* Michael R. Dougherty & Amber Sprenger, *The Influence of Improper Sets of Information on Judgment: How Irrelevant Information Can Bias Judged Probability*, 135 J. EXPERIMENTAL PSYCHOL.: GEN. 262, 262 (2006).

38. *See* John Markoff, *Armies of Expensive Lawyers, Replaced by Cheaper Software*, N.Y. TIMES, Mar. 4, 2011, <http://www.nytimes.com/2011/03/05/science/05legal.html>. Thanks to Drausin Wulsin for bringing my attention to this article.

39. *United States v. Williams*, No. 09-5256, 2011 U.S. App. LEXIS 9550, at \*26-27 (6th Cir. May 11, 2011) (Thapar, J., concurring).

*A. Watson the Judge*

Could Watson perform better than judges at the tasks of statutory interpretation? Each of the three elements of new textual interpretation—premise, process, and reasoning—point toward the possibility of Watson outperforming new textualist judges at their own game.

First, computers support new textualists' *premise* by offering a mechanical way of determining the "ordinary meaning" of a statute. According to *Merriam-Webster's Collegiate Dictionary*, "ordinary" means "of a kind to be expected in the normal order of events; routine; usual."<sup>40</sup> The common factor in each part of the definition is frequency; given a set of circumstances, the ordinary outcome is the outcome that occurs more often than other possible outcomes. Humans are flawed textualists because they have only one frame of reference: their own "ordinary" experience. Any computer is better equipped to identify the frequency with which a particular phrase occurs in common parlance.

Take a famous example: in *Muscarello v. United States*,<sup>41</sup> the Supreme Court debated the meaning of the phrase "carries a firearm." The majority argued that the ordinary meaning of carrying a gun included transporting it in a vehicle.<sup>42</sup> The dissent disagreed, arguing that "carry" required holding a gun on one's person.<sup>43</sup> The two sides marshaled a vast array of evidence from the public domain to demonstrate that their interpretation was the most ordinary, including dictionaries, news articles, and even the Bible. Watson could have saved the Court's law clerks a great deal of trouble. The computer would have been able to calculate how frequently the terms "carry" and "vehicle" (or their synonyms) appear together versus "carry" and "person" (or their synonyms). Thus, in at least one sense Watson is better at textualist interpretation than humans—he can not only identify ordinary meanings but can tell us just how ordinary a particular meaning is!

Watson's superior recall is particularly important given the historical nature of statutes, meanings of which can change over time. Justice Scalia, for example, has suggested that absolute immunity for prosecutors did not exist at common law.<sup>44</sup> A well-informed Watson could report back in a matter of minutes as to the likelihood that this was true. Watson even may be able to help decipher antiquated meanings on which there is no modern

---

40. MERRIAM-WEBSTER'S COLLEGIATE DICTIONARY 873 (11th ed. 2004).

41. 524 U.S. 125 (1998).

42. *Id.* at 126-27.

43. *Id.* at 140 (Ginsburg, J., dissenting).

44. *Burns v. Reed*, 500 U.S. 478, 496-97 (1990) (Scalia, J., dissenting).

expertise—such as common law phrases no longer used today—by looking at the context in which such phrases were used.

This raises a second Watsonian virtue: his *process* of interpretation. Most computers merely isolate instances where identical words appear most closely to one another. Watson’s algorithms go a step further by distinguishing which connotation of a particular word is intended based on the particular context. Watson might not only look for words elsewhere in the statute, but could also draw from other words *not* in the statute to provide additional interpretative context. In the *Muscarello* example, there was at least one contextually-appropriate usage of “carry” that was not uncovered by either party in the litigation: whether state “carry” gun laws (for example, “open carry” and “concealed carry” gun laws) apply to vehicles.<sup>45</sup> Watson could have estimated the frequency with which each connotation arises—including the state law use of “carry” not considered by the actual parties—to determine whether “carry” ordinarily encompasses transportation in vehicles.

Finally, and perhaps most importantly, Watson’s reasoning is more systematic than humans’ reasoning. Inasmuch as he makes errors, these errors are randomly distributed. His mistakes are not skewed due to political preferences, personal relationships, or other sources of human prejudice. Watson by design avoids the ideological bias of judges—which textualists so deeply fear—because, of course, he does not have any ideology of his own. These arguments are summarized in Figure 1.

---

45. I am indebted to Farah Peterson for this point. See Farah Peterson, *A Dialectical Federalism for Statutory Interpretation* (May 26, 2011) (unpublished seminar paper) (manuscript at 45-50) (on file with author).

Figure 1  
WATSON VERSUS NEW TEXTUALISM

	Elements of New Textualism	Elements of Watson	Watson's Advantage
<b>Premise</b>	"Ordinary Meaning"	Knowledge Breadth	Can Determine Frequencies of Meanings
<b>Process</b>	Evaluating Context	Relationships Among Words	Can Consider Unrelated Contexts for Clues
<b>Reasoning</b>	Avoiding Bias	Probabilities	Errors Are Random

### *B. Watson's Limitations as a New Textualist*

Despite these advantages, computers are unlikely to replace judges anytime soon. For one thing, Watson still makes mistakes at critical times. Perhaps the most amusing occurred in the very last *Jeopardy!* round in the competition. The Final Jeopardy category was "U.S. Cities," and the answer was the following: "Its largest airport is named for a World War II hero; its second largest, for a World War II battle." While both Jennings and Rutter correctly provided the question "What is Chicago?" Watson responded, "What is Toronto?????"<sup>46</sup> Given that the only city named Toronto with any commercial airport is not in the United States, but in Canada, this was a baffling response.

The Toronto incident highlights that Watson cannot filter away such absurd responses on his own. Without a human to assist him, serious errors may remain. To be fair to Watson, the question marks indicate he was highly unsure about his response to the "Toronto" question; he was forced to answer the question in Final Jeopardy and wagered a low amount as a result.<sup>47</sup> But this quantified uncertainty may not be useful when Watson attempts textualist interpretation. If Watson is uncertain about the "ordinary meaning" of a statute, he will not be able to refuse to buzz in.

46. Samara Lynn, *Watson Trounces, Then Trips on Jeopardy!*, PCMAG.COM, Feb. 16, 2011, <http://www.pcmag.com/article2/0,2817,2380429,00.asp>.

47. See Markoff, *supra* note 30.

When Watson can find no clear ordinary meaning, what should he (or a judge) do then?

This suggests that the most serious critique for a Watson-led textualism is not practical but principled: at least in the tough cases, judging *should* contain normative as well as objective inputs. Employing Watson for statutory interpretation requires an important choice between allowing judicial decisions with random error but occasionally absurd results or allowing decisions with nonrandom, biased error. Watson could achieve the new textualists' stated goal of determining ordinary meaning—with a dash of random error. But he could never decide, for example, that an outcome is normatively absurd. According to his computational frame of reference, any answer his algorithm spits out is the most likely accurate meaning. Do new textualists really want judicial decisions to be made based only on the frequency with which a meaning appears in Watson's memory, especially when his certainty is low? I expect not.

### *C. Watson Assisting Textualists*

As IBM brings Watson's DeepQA technology to the medical community, Watson's creators are not proposing that his algorithms could replace doctors in their entirety. More appropriately, they suggest that Watson could aid doctors in doing their job better. This also may be the most appropriate role for Watson in the judicial sphere. A Watson-type tool could bring the advantages of computer-based analysis to statutory interpretation without sacrificing the normative discretion which allows humans to "get it right" in ways that computers cannot.

One can imagine a tool into which users could input short phrases from statutes. The tool, powered by DeepQA technology, would then output the ordinary meaning of the phrase based on frequency calculations. Such a technology would create a presumption of ordinary meaning that judges would become (informally) bound to refute if they wished to stray from such meaning. Circumstances in which they might stray might include (1) a close call where two 'ordinary meanings' score highly; (2) an analytically dubious result accompanied by a low level of confidence (i.e., the Toronto example above); or (3) a normatively absurd result produced by the effects of the ordinary meaning, such as an excessively punitive result. This is similar to the way that judges already treat government agencies: in most cases giving them deference.<sup>48</sup>

---

48. *Cf. Nat'l Cable & Telecomms. Ass'n v. Brand X Internet Servs.*, 545 U.S. 967, 980 (2005) ("If a statute is ambiguous, and if the implementing agency's construction is

How might judges employ such a technology? One possibility is that it could serve the function of a law clerk and conduct basic research upon which judges can construct their opinions. A second possibility is that it could become a resource of the Federal Judicial Center; officials at each courthouse could get trained in the technology. Third, Watson might function usefully as a tool for the private sector. Lexis and Westlaw might purchase the rights to the technology, providing firms, universities, and judges alike with the ability to determine their own “ordinary meanings.” By providing more definitive meanings, Watson could eventually reduce litigation—if all parties agree to turn their fate over to the hands of a computer.

### CONCLUSION

Watson achieved a great victory for computational “thinking” over human “thinking.” But he cannot yet make the normative decisions that ethical judging requires. What Watson already can do for judges is to provide a baseline against which to evaluate their own interpretations of “ordinary meaning.” Watson will not stop bias from creeping into judicial decisionmaking—but his contributions to statutory interpretation are nevertheless far from trivial.

*Betsy Cooper is the Executive Bluebook Editor of The Yale Law Journal and a member of the Yale Law School Class of 2012. She received her DPhil in Politics from the University of Oxford in 2009. The author would like to thank Aaron Barkhouse for inspiring this Essay and Professor William Eskridge, Arpit Garg, Daniel Hemel, Nick Hoy, and The Yale Law Journal Volume 121 team for their helpful feedback.*

Preferred Citation: Betsy Cooper, *Judges in Jeopardy!: Could IBM’s Watson Beat Courts at Their Own Game?*, 121 YALE L.J. ONLINE 87 (2011), [www.yalelawjournal.org/2011/08/23/cooper.html](http://www.yalelawjournal.org/2011/08/23/cooper.html).

---

reasonable, *Chevron* requires a federal court to accept the agency’s construction of the statute, even if the agency’s reading differs from what the court believes is the best statutory interpretation.”); *Chevron U.S.A. Inc. v. Natural Res. Def. Council, Inc.*, 467 U.S. 837, 844 (1984) (establishing that administrative regulations interpreting ambiguous language in a statute “are given controlling weight unless they are arbitrary, capricious, or manifestly contrary to the statute”).