What Did Watson the Computer Do?

By STANLEY FISH

Last week brought the story of a fourth grader who was suspended from school for having placed a "kick me please" post-it on the backside of a classmate. The punishment, The New York Post reported, was delivered by the school principal who said in a note to the boy's parents, "This incident is in violation of the Discipline Code and is classified as infraction A37 — engaging in bullying behavior."

The code was part of a "zero tolerance" policy adopted by the New York City school system and its application in this case tells us what zero tolerance means. It means no deviation from a precisely and narrowly formulated rule. "Narrowly formulated" is redundant because a rule is supposed to be narrow in a way that leaves very little if any latitude for interpretation. A rule does not encourage one to ask if the fourth grader's act was bullying or teasing; if it fits the physical description of bullying, then bullying or infraction A37 is what it is. A rule is an "all or nothing" proposition; it lays down the law and admits of no exceptions that might be claimed on the basis of all the considerations it purposely excludes. When strong First-Amendment types encounter an argument for regulating speech in certain circumstances, they respond by asking (rhetorically), "What part of 'make no law' don't you understand ?" That's what's good about a rule; it stands as a bulwark against the instability and unpredictability that come along with making decisions case-by-case.

That's also why people — human beings — have trouble keeping to the rules. Human beings are always thinking, "Yes, I know the rule, but surely those who crafted it would agree that in the situation I now face, it should be relaxed" or "I know the rules of this game but if I obey them slavishly I'm likely to lose, so why don't I bend them creatively and see if I can get away with it." And then there are the familiar conundrums: The rule is no jaywalking, but only by jaywalking will I have a chance of saving the dog that is about to be hit by a car. Or I know the rule is don't lie, but telling the truth now might endanger an innocent person. Or the rule is don't travel at a speed above 25 miles per hour in the city limits, but my wife is about to give birth.

The inability or unwillingness of human beings to follow rules or be content with their guidance is not a weakness but a strength.

The rule we find unhelpful (or inconvenient) was written with a particular set of circumstances in mind, but circumstances not contemplated by the rule-makers will always turn up. When they do, rule aficionados will say, "Well, we can emend it and make it more supple by adding to the circumstances it covers. But you can never add enough; the proliferation of circumstances always outruns the efforts to take account of them, and after a while you've reached the point when every situation will require a rewriting of the rule, which means that there will no longer be a rule at all.

If you have followed the argument so far, you will have anticipated its next turn, which is to say that the inability or unwillingness of human beings to follow rules or be content with their guidance is not a weakness but a strength; it is the strength of being being able to adjust when the rules have nothing helpful to say or produce absurd results in a situation the rule-makers did not anticipate. Only a fool will persist in adhering to a rule or set of directives when its application is clearly counter-intuitive and even disastrous. Those who are not fools will think that this is a new game — a new situation — and it calls for new strategies, different calculations of what will and will not work.

The computer I am writing this column on is a fool. It has a program that directs it to finish words

before I do by “consulting” a data base of words I have used that begin with the letters I have already typed. “Consulting” is in quotation marks because the computer isn’t doing anything that requires intelligence as opposed to calculation; it is sorting through data and matching the data it has stored with the data of my initially chosen letters. It is almost always wrong because its procedures do not track my practice. I am not self-consciously generating a pattern of statistical frequencies. I am producing words that have been chosen because they contribute to the realization of a governing idea or a compositional plan. In fact, to say that the computer is wrong is to give it more credit than it deserves; for right and wrong are not what it does; what it does is count (faster than I or anyone else could) and match. What it doesn’t do is begin with an awareness of a situation and an overall purpose and look around for likely courses of action within that awareness. That is because, as the philosopher Hubert Dreyfus explained almost 40 years ago, a “computer is not in a situation” (“What Computers Can’t Do”); it has no holistic sense of context and no ability to to survey possibilities from a contextual perspective; it doesn’t begin with what Wittgenstein terms a “form of life,” but must build up a form of life, a world, from the only thing it has and is, “bits of context-free, completely determinate data.” And since the data, no matter how large in quantity, can never add up to a context and will always remain discrete bits, the world can never be built.

What computers can’t do, we don’t have to do because the worlds we live in are already built; we don’t walk around putting discrete items together until they add up to a context; we walk around with a contextual sense — a sense of where we are and what’s at stake and what our resources are — already in place; we inhabit worldly spaces already organized by purposes, projects and expectations. The computer inhabits nothing and has no purposes and because it has no purposes it cannot alter its present (wholly predetermined) “behavior” when it fails to advance the purposes it doesn’t have. When as human beings we determine that “the data coming in make no sense” relative to what we want to do, we can, Dreyfus explains “try a new total hypothesis,” begin afresh. A computer, in contrast, “could at best be programmed to try out a series of hypotheses to see which best fit the fixed data.”

That’s what Watson — the I.B.M.-built computer that won a game of “Jeopardy” last week over two human opponents — does. It’s just a bigger and fancier version of my laptop’s totally annoying program. It decomposes the question put to it into discrete bits of data and then searches its vast data base for statistically frequent combinations of the bits it is working with. The achievement is impressive but it is a wholly formal achievement that involves no knowledge (the computer doesn’t know anything in the relevant sense of “know”); and it does not come within a million miles of replicating the achievements of everyday human thought.

Watson’s builders know this; when they are interviewed they are careful to stay away from claims that their creation simulates human mental processes (although they also murmur something about future hopes). But those in charge of the artificial intelligence hype are not so careful and they delight in exciting us and frightening us with the fiction of a machine that can think. It’s great theater, or in Watson’s case, great television, but that’s all it is.