

# **Wason's Puzzle and Real Problems**

by Howard Margolis

Irving B. Harris Graduate School Of Public Policy Studies  
University of Chicago  
1155 East 60<sup>th</sup> Street  
Chicago, Illinois 60637

October 2002

Abstract: Two very strange results from variants of the much-studied Wason selection task point to an interpretation which has substantial implications for important choices in the world.

(Prepared for AIBS conference on behavioral economics, Great Barrington MA, 2002)

## 10.14.02

### **Wason's Puzzle and Real Problems**

by Howard Margolis

(paper for AIBS conference on behavioral economics, Great Barrington MA, 2002)

Abstract: Two very strange results from variants of the much-studied Wason selection task point to an interpretation which has substantial implications for important choices in the world.

#### **1. Introduction**

Intuition—what we see as right prior to thinking about why it is right—must be rooted in experience. In substantial measure that seems to be rooted in the fossilized experience of our Darwinian heritage. And on evolutionary logic we ought to expect that entrenched propensities mostly perform well (or what could account for their being entrenched?) But we also should expect that what had been long-ago entrenched will not always be reliable. In unfamiliar contexts that somehow look like the kind of contexts which promoted entrenchment, we can have sharp intuitions that feel like insights but turn out to be illusions. Psychologists have accumulated a kind of zoo of specimens of cognitive illusions. In turn this has produced what by now is a long-running controversy over how important illusory intuitions might be. Are they really much more than parlor game amusements and grist for undergraduate psychology experiments?

Gerd Gigerenzer and his colleagues have done admirable work on the skeptical side of this debate. But my own inclinations plainly line me up on the side of Kahneman & Tversky and others who think the illusions are extremely important. No one thinks intuition is mostly illusory. We can reasonably suppose that illusion is only a small part of what makes intuition important. But a dash of arsenic in a glass of fine wine is also only a small part of the whole.

Think of a horizontal "V". In the middle we can imagine the zone of familiar experience, where we have plenty of opportunity to learn what works. But at the narrow end we have the impoverished environment of artificial puzzles, some of which (like the elaborately discussed Monty Hall problem) lead to almost irresistible but pitifully unsound intuitions. And at the wide end we have cases out of scale or beyond the range of familiar experience, where our ordinarily reliable intuitions can again lead us seriously astray. And then -- as voters, jurors, investors (to mention the most obvious examples) -- we may be misled by confident but seriously distorted intuition.

In a new book I use such ideas to give a new account of the Scientific Revolution, which can

reasonably be seen as the most impressive of all "wide-end" cases.<sup>1</sup> For cutting edge science is by definition dealing with what is beyond what we know from experience. But at the very beginning of modern science almost everything was cutting-edge.

Table 1, from the opening pages of the book, shows what an astonishing burst of discovery occurred in the years immediately bracketing 1600. The book is about how to account for that astonishing burst. And here a critical clue comes from noticing that although some of these c. 1600 discoveries could not have been made any earlier than they were (because they required newly available data or a newly available instrument), most required nothing that was not available 2000 years earlier.

---

<sup>1</sup> Margolis 2002. The book's Epilog argues the relevance to contemporary social choice. This paper, starting from the opposite end of the spectrum, extends that argument. As with physical habits, difficulties turning on cognitive habits can occur either when an entrenched habit is inappropriately triggered, or when a novel habit must be learned. For the Scientific Revolution, the key turn appears to have been the emergence of a novel habit of mind which prompts "around the corner" inquiry. I show what seems to me very strong evidence tying this novelty to the Copernican commitments of all four of the men who jointly produced all the discoveries in Table 1. But in ordinary social contexts, as in narrow-end experiments, such as the Wason puzzle discussed here, novel habits can hardly appear. It is the consequences of inappropriate triggering of entrenched habits that will concern us.

## Table I-1 Notable Scientific Discoveries

Made c. 1600	Made in the Previous 14 Centuries
*Distinction between electricity and magnetism	
*Law of free fall	
*Galilean inertia	
*Earth is a magnet	
*Theory of lenses	
*Laws of planetary motion	
*Various discoveries with the telescope (the Moon is Earthlike, with mountains; the Sun exhibits spots and rotates; Jupiter has moons; etc.)	
*Laws of hydrostatic pressure	
*Synchronicity of the pendulum	

---

Why did we have to wait 2000 years? But since we did have to wait 2000 years, it seemed pretty certain that cognitive oddities must be part of this story. The actors who missed these discoveries for 2000 years were not undergraduates going through the motions of solving puzzles they do not really see as important. Nor will anyone suppose that around 1600 we just abruptly got smarter. And it also seems completely unlikely that *all* consequences of whatever delayed most of the discoveries in Table 1 for 2000 years evaporated in 1600.

### 2. Wason's Puzzle

But turn for now from this very grand “wide end” topic to a pair of curious results at the other (narrow) end of the spectrum. Here we are indeed only toying with parlor game questions. But these little puzzles have the advantage of being easy to vary and repeat with new subjects. So we can sometimes find new responses that provide clues into what goes on inside our heads.

During the 1960's an English psychologist (Peter Wason) spent a visiting term at Harvard, working with Jerome Bruner on the early stages of what has come to be called “the cognitive revolution”. Wason was interested in cognitive illusions and exceedingly clever at devising interesting examples. During that visit he formulated what has turned out to be the most extensively studied puzzle in the history of psychology.

Here is a version of this famous “selection task”. Cards are labeled “A” or “D” on one side

and "2" or "3" on the other. A rule says that "if A then 2". Subjects see an array of four of the cards, two letter-side up (showing "A" and "D") and two number-side up (showing "2" and "3"). A subject must decide which cards need to be turned over to check whether the rule has been violated. If you are not already familiar with the puzzle, you might try to answer before reading further.

The common responses are "A & 2", or "A" alone. A large majority typically give one of these responses. And except for psychologists and others familiar with the puzzle this usually holds for professors, lawyers and so on about as much as for undergraduates. The correct response is "A & 3".<sup>2</sup>

Since Wason introduced the task, a vast amount of testing of variants has produced a consensus that what the illusion shows is that human beings are generally incapable of performing the "modus tollens" inference required to give the correct response, unless concrete content is provided.<sup>3</sup> But in an unfamiliar or abstract context like the classic version of Wason considered here, we mostly can't... or so several decades of psychology students have been taught.<sup>4</sup>

### 3. A Strange Result

But the claim that inability to handle modus tollens explains Wason is thrown into doubt by a very odd feature of the Wason task, noticed early on by Wason himself but then almost completely ignored by everyone (even Wason himself!). This is the performance of subjects when the cards, which nearly everyone gets right, are removed. Wason's own 1970 discussion (with Johnson-Laird) of what he called a "reduced array" perhaps blurred the essential point by using many cards in a more complicated version of the task.

With the standard 4-card version, anyone can see that there are two easy cards which hardly ever produce errors: "A", which is rarely missed, and "D", which is rarely chosen; and two hard

---

<sup>2</sup> The "A" might have a "3" on its down-side, and the "3" might have an "A". Either case violates the rule. But the rule does not restrict what is on the reverse side of a "2" or of a "D". So turning either of these cards is unnecessary: whatever you find will not violate the rule.

<sup>3</sup> For "if A then 2", the modus tollens inference is: "so if not 2, then not A", as in "if he's hungry, he will eat, but since he doesn't eat, I infer he is not hungry." Here since "A" should have "2" on its reverse side, "3" should *not* have "A" on its reverse side. So you need to turn the "3" to check for that possible violation.

<sup>4</sup> The broad consensus is not unanimous. An early version of this paper (Margolis 2000) generated a set of comments from people who had also found experimental evidence that inability to perform "modus tollens" was not really the explanation of the illusion. See Margolis 2001.

cards: "3" and "2", which supply nearly all the errors. Consequently, we would hardly expect a big change in performance if we eliminate the easy cards, which almost no one misjudges anyway. This apparently seemed so obvious to Wason himself that he did not bother to try the very simple reduced array discussed here: he seems to have taken it for granted that it would have no interesting effect.

But that turns out to be very wrong. If subjects in fact are shown only the two hard cards, leaving out the two easy cards, we find that leaving out the easy cards seems to correct the errors on the hard cards!

If a trial is run showing half the subjects all four cards (A,D,2,3), but showing the other half only the hard cards (2,3), my experience with classroom exercises is that only the 4-card version returns the usual overwhelmingly wrong responses. Those given the reduced array return a clear majority of correct responses!

What can possibly account for this large improvement, related to merely removing the two cards that are ordinarily, and quite effortlessly, judged correctly anyway?

#### 4. Seeing Categories

The explanation I've proposed (Margolis 1987) may seem as unlikely as the result it explains. But you can easily do this very simple experiment and satisfy yourself about whether you do get the bizarre result. And I will follow an explanation of what happens with a very solidly confirmed second variant of Wason which is equally surprising in another way, but which makes sense if the "reduced array" explanation is right.

Almost no one has any difficulty understanding the plain meaning of the Wason question. The question is certainly about the four particular cards shown, and if asked, people easily see that. And people also turn out to be competent to follow a "modus tollens" inference even if they apparently have just failed to use it. Although most of us prove vulnerable to Wason's illusion, it is rare that anyone has any trouble at all understanding why "A & 3" is the right answer once it is pointed out.

But a key to what is happening can be found in noticing that the usual "A & 2" or "A only" responses (both wrong) make sense if somehow subjects responds *as if* the question were about which *categories* of cards should be examined... treating, for example, the "D" card as representing any cards with a "D", rather than seeing it as the particular card shown with a "D" on its upside.

If subjects are tacitly (certainly not consciously) treating the cards as representing categories

of cards (not as the particular cards Wason asks about), then the *salient* correct response to that incorrect perception is either "A & 2" when "if/then" is read as "if and only if" (iff), or "A" alone when "if/then" is read as "if but not necessarily only if" (if): just the pair of responses we do most often see.<sup>5</sup>

In Margolis 1987 (pp. 151-2) I suggested how the illusory "categories" reading can arise from entrenched expectations. We ordinarily first have to choose what sort of approach to take (here what categories are relevant) and only then deal with details of how to do it. In the absence of sufficiently rich context (eg., in the abstract Wason context), we apparently tend to fall into seeing first what we usually do first. In this impoverished context, that tendency turns out to override what the words in fact tell the subject to do, as a person used to turning an upper lock before a lower may find himself reaching for the upper lock even after it has become non-functional.

But consider what happens when we eliminate the two easy cards ("A" & "D"). If the illusory "category" perception of the task is tacitly guiding intuition, then the proper response (to that illusory perception) is "2 & 3" for the "iff" reading of "if/then", or "3" only for the "if" reading. So now a correct response for the "categories" misreading *coincides* with a correct response for the intended reading. In particular, it no longer misses the "3". The illusion need not be paradoxically cured by removing the easy cards. But now it becomes invisible!

And a norm of language favors "3" alone (not "2 & 3") with the reduced array. Other than for rhetorical purposes, we do not ask questions with obvious answers. This favors the "if" (rather than "iff") reading, which gives the solver a bit more to think about: favoring "3" alone over both "2 & 3".

In sum, eliminating the easy cards does not make the problem actually easier. All that is actually happening is that the salient "correct" (using modus tollens) response to the illusory perception is no longer available. So subjects quite readily (using -- not missing -- modus tollens) give the response that is available with the reduced array. But now the only response that fits includes the card usually missed. And then the pragmatics of language usually also eliminates the unnecessary (though not really wrong) "2" choice.<sup>6</sup>

---

<sup>5</sup> I need to stress that these are the salient responses, made so by being the cards mentioned in the question ("A" and "2"). But three more pairs in addition to "A & 2" would also be correct for a "categories" response to the "iff" reading: "A & D", "2 & 3", "D & 3". Any of these choices will locate all violations (A/3 or D/2 cards). And for the "if" reading "3" as well as "A" would be correct, finding any A/3 cards.

<sup>6</sup> If subjects did usually choose "2&3" rather than "3" alone, that would not really be a clear error. Outside of formal logic, whether "if, then" might mean "if and only if" (rather than "if, but not necessarily only if") is usually

## 5. The “two cards” experiment

Another strange results lends strong support to this account. A secondary point of the “categories” account suggested why with the reduced array subjects not only turn the "3" but usually no longer turn the unnecessary "2". The point was that the too-easy character of the question -- when read to require checking all the cards available -- nudges subjects toward an "if" (rather than "iff") response. Anyone familiar with the ways of stage magicians will know that, when neither choice is clearly marked, subtle nudges indeed can have large effects. And the Wason context itself provides a striking example.

Griggs (1990) confirmed a surprising effect that I had published in Margolis 1987. I had used a variation of the Wason task intended to force responses especially heavily towards "A & 2". But then a quite trivial variation in wording turned out to shift responses very heavily to the otherwise almost never seen response of "D & 3"! But this "D & 3" response (in logical notation, the "not-P, not-Q" response) is in fact yet another correct response to the illusory perception of the cards as labels for categories (recall again note 5).

The change in wording was merely from: "Circle two cards to turn over to check whether the rule has been violated" (eliciting exceptionally heavy "A & 2" responses) to: "Figure out which two cards could violate the rule, and circle them" (which changes the dominant response from "A & 2" to "D & 3".)

That "D & 3" is rarely seen in the dozens of published Wason experiments shows the ordinarily marked salience of "A" and "2", which are explicitly mentioned in the rule. The two responses are equally “correct”.<sup>7</sup> But the seemingly innocuous shift to the second version of the instruction (we can see from this reliably replicable result) makes the cards *not* mentioned in the rule salient. Again the response is a response *using* (not missing) modus tollens, but to the misperceived “categories” sense of the question.

## 6. Social Consequences

Now consider possible social implications of these odd results. Wason's puzzle is certainly trivial. But what we can learn from it does not seem to me trivial.

---

left implicit, since in context this hardly ever causes difficulty. But here the context is so thin that a person could plausibly read "if" either way.

<sup>7</sup> “Correct”, that is, as a proper response to the illusory *categories* misreading of the task.

On the account now sketched what governs Wason is a cognitive illusion at the stage of recognizing the intended task, not incompetence at handling the intended task due to an inability to handle elementary logic. And it is an important feature of the account that the illusory perception of the task is *unconscious*. That is clear. For if asked, no one says (or once the error is pointed out, recalls) interpreting the question as about categories.

On the other hand, the usual claim that Wason shows an inability of ordinarily competent people to handle modus tollens has always warranted more suspicion than it has received. Anyone who listens to their children will hear them quite readily make what are functional equivalents of modus tollens inferences. And that can hardly be surprising, since the world provides us all with endless occasions to make such inferences. (If I picked my keys off the desk, they would now be in my pocket. My keys are not in my pocket. So they are probably still on my desk.) Modus tollens inferences, rather than being beyond normal competence other than in particular favorable contexts, are routinely used by anyone of normal competence. So however surprising the interpretation of Wason urged here, at least it does not turn on a claimed inability of subjects to do what we all are routinely capable of doing in everyday life.

And that has consequences for the significance of the illusion.

For if the correct explanation of Wason were indeed that people are not competent to manage modus tollens inferences in an unfamiliar context, then it is hard to see large consequences from that. Serious questions get attention and discussion. And we can see in Wason that once the missed inference is pointed out, people quickly get the point. So the flaw in reasoning would be noticed, and once noticed it would not be defended. But the Wason variants just reviewed imply that the illusion does not in fact turn on missing a logical step. The difficulty has an essentially different character. It is that in a context outside a person's ordinary experience, intuition may respond, knee-jerk fashion, to what subliminally "looks like" the appropriate context but isn't, and even in the face of simple language that you would expect would make the correct recognition trivial.

That possibility poses a risk in situations like those jurors, investors and voters (for examples) often face. For their choices will not have the naked simplicity of a puzzle question like Wason. There will not be the starkly simple way to show jurors, investors and voters when an intuition is mistaken that you see for the Wason explanation in note 3. Even if, as in Wason, judgment is unsound in terms of a person's own standards of how judgment should work, the origin of the

---

difficulty may be quite out of sight. Further, as in Wason, the difficulty can be not only out of sight of the person making the judgment but hard to detect also by someone observing the judgment. Then it would be naïve indeed to suppose that misunderstandings can be resolved by some starkly simple argument.

And in contrast to puzzles like Wason, in larger contexts whatever cognitive difficulties are present will also be compounded by sympathy, passions and interests. So in yet another way, and hardly an unimportant way, persuading someone that his judgment turns on a misperception which he cannot see, and that even neutral parties cannot easily see, will be vastly more difficult than pointing to a logical slip which becomes undeniable once it is pointed out.

This is already a large menu of difficulties. But there is more.

Social actors are reluctant to bluntly attack a judgment (other than one coming from an irreconcilable adversary) which others feels strongly is right. We often avoid such confrontations out of simple politeness. We also do it for strategic reasons. We want to be trusted and listened to on other matters where the stakes may seem more important or the prospects of success more promising. Beyond all that, we are all influenced by how our social peers are seeing things. And the cognitive effects most relevant in social contexts are hardly likely to be narrowly idiosyncratic. Persuasion often has to proceed in the face of a *widely-shared* sense that "everyone knows" the opposite is correct. The prospects of making persuasive even what in hindsight will seem a terrifically good case, when it runs against popular intuition, is rarely very good. The task is not always hopeless, especially over time, but if it is often difficult enough to discourage even much effort.

So are cognitive illusions likely to occur and be hard to correct beyond the narrow contexts of psychology experiments? I think that is inevitable, so that it is important that we come to understand these illusions. Careful thought about simple puzzles like Wason can help illuminate what happens in vastly more consequential realms. Perhaps this can now even be regarded as almost a mainstream view with the blessings of Stockholm on Danny Kahneman.

## References

Griggs, R. (1990) "Instructional effects on responses in Wason's selection task." *British Journal of Psychology* 81:197-204

Johnson-Laird, P.N. and Wason, P.C. (1970) "Insight into a logical relation." *Quarterly Journal of Experimental Psychology* 22:49-61.

Margolis, H. (1987) *Patterns, Thinking and Cognition*. University of Chicago Press

\_\_\_\_\_ (2000) Wason's Selection Task With Reduced Array. *Psychology* 11(005)

\_\_\_\_\_ (2001) More on Modus Tollens and the Wason Task. *Psychology* 12(010)

\_\_\_\_\_ (2002) *It Started with Copernicus*. McGraw-Hill.

Wason, P.C. (1983) Realism and rationality in the selection task. In Evans, J.S.T.B. (Ed.) "Thinking and Reasoning" Routledge & Kegan Paul.