Notes for Week 13 of *Confirmation*

12/05/07

Branden Fitelson

1 The Wason Selection Task

1.1 The Task Itself

Here is Wason's "description" of the "selection task", in some detail:

> ...given the sentence: Every card which has a D on one side has a 3 on the other side (and knowledge that each card has a letter on one side and a number on the other side), together with four cards showing respectively D, K, 3, 7, hardly any individuals make the correct choice of cards to turn over (D and 7) in order to determine the truth of the sentence. This problem is called the "selection task" and the conditional sentence is called "the rule". The rule has the logical form, "if p then q" where p refers to the stimulus mentioned in the antecedent (D); $\bar{p}$, i.e. not p, refers to the stimulus which negates it (K); q refers to the stimulus mentioned in the consequent (3); and $\bar{q}$, i.e. not q, refers to the stimulus which negates it (7). In order to solve the problem it is necessary and sufficient to choose p and q, since if these stimuli were to occur on the same card the rule would be false but otherwise true. The combined results of four experiments ...show that the subjects are dominated by verification rather than falsification. On the whole, they failed to select $\bar{q}$, which could have falsified the rule (and they did select q, which could not have falsified it although this latter error is much less prevalent).

Wason's description is slightly odd. Here's how I would describe it. Assume the subjects are told (antecedently) that they are going to be testing a hypothesis involving four cards. Each card has one letter on one side and one number on the other side. They will be shown the four cards (face down), and they will be asked to turn over one or more of the cards, with an eye toward determining whether the following is true (where they are to turn over *the fewest cards possible* for that are sufficient to make this determination):

(H) Every card that has a "D" on one side has a "3" on the other side.

In our terminology, the subjects will be generating some *evidence* (E) regarding H, and their goal is to generate an E which determines whether or not H is true (in such a way that E involves *the fewest cards possible* for this task). Wason says that the "rule" (H) has logical form "if p then q". This isn't quite accurate. I would say that the logical form of the sentence is "$(\forall x)(Dx \supset Tx)$". Nonetheless, one could think of a conditional version of H, which might look like the following:

(H') If a card has a "D" on one side, then it has a "3" on the other side.

My sense is that this version also has very similar response patterns. Nonetheless, I will use H rather than H', since it ties in more nicely with Hempel's raven paradox. However, I will try to be more (formally) precise about the descriptions involved. We have four cards: a, b, c, d, and I will use the following predicates: $Dx \overset{\text{def}}{=} x$ has a "D" on one side, $Tx \overset{\text{def}}{=} x$ has a "3" on one side, $Kx \overset{\text{def}}{=} x$ has a "K" on one side, and $Sx \overset{\text{def}}{=} x$ has a "7" on one side. The specific way in which the cards are shown to the agents (which is also part of their background evidence, let us assume) is as follows: D, K, 3, 7. That is, they know (antecedently) that there are four cards (a, b, c, d) in the domain of discourse, and that $Da \,\&\, Kb \,\&\, Tc \,\&\, Sd$. Then, they are asked to turn over some of the cards (and the fewest number possible for this purpose), so as to determine whether H is true (on said domain of discourse — ambiguities about the domain of discourse will become crucial).

1.2 The Empirical Patterns of Response

Tasks such as these have rather robust response patterns from subjects. Wason's observed patterns were:

	Strategy	Proportion who chose strategy
(S_1)	Turn over a and c.	59/128
(S_2)	Turn over a.	42/128
(S_3)	Turn over a, c, and d.	9/128
(S_4)	Turn over a and d.	5/128
	other	13/128

The *minimal*-card # strategy which will *falsify* H *if* it's *false* and *verify* H *if* it's *true* (relative to the background evidence given), is strategy S_4: turning over a and d. This is because we know antecedently that the only possible counterexamples to H are a and d. One might try to use confirmation theory to "explain" why subjects opt for S_1 most of the time (and the other patterns). At least, several authors have tried to do this.

2 Wason Meets Hempel?

2.1 Humberstone, Nickerson, and Hempel

The first person to suggest a clear parallel between Wason and the raven paradox was Humberstone. Nickerson has a more recent psychology paper (on website), but, sadly, he doesn't give Humberstone any credit. This is too bad, but Nickerson's analysis goes far beyond Humberstone's (Nickerson gives a very nice Bayesian confirmation-theoretic analysis, which I will discuss below). Here is Humberstone's basic idea:

> ...imagine that the experimenter's cards have been thoroughly inspected by the subjects, and found each to have the name of a bird on one side and of a (not necessarily chromatic) colour on the other. Four are drawn at random and placed on the table, showing "Raven", "Swan", "White" and "Black". The hypothesis to be tested is then what we shall call $H^\star$:
>
> ($H^\star$) If a card has "Raven" on one side, then it has "Black" on the other.

In Humberstone's story, he talks about *two distinct renditions* of the task. We could take the task to be just like the Wason task, where *the domain of discourse is just the four cards on the table.* Or, we could take the domain of discourse to be *all the cards in the deck.* [This ambiguity about the domain of discourse is *crucial* to the analogy, and also to Nickerson's Bayesian analysis. But, unfortunatley, Nickerson doesn't give credit to Humberstone for this crucial observation.] The second rendition is more analogous to Hempel's paradox, since there is no way to verify Hempel's hypothesis with a finite number of observations, and there is no way to verify the hypothesis about all the cards in the deck just by inspecting some of the four cards on the table. Moreover, the first rendition leaves very little room for confrimation-theoretically explaining (in a *rationally compelling* way) why subjects might exhibit the patterns they do (there, *the* answer seems *clear*). But, if we allow ourselves to think about the second rendition, there is room for Bayesian confirmation theory to say some interesting things about the Wason task (as Nickerson does very well in his paper). Before I get into that, I want to discuss a couple of possible disanalogies between Hempel and Wason. The first disanalogy applies to both renditions, the second disanalogy applies only to the second rendition.

First, recall that Hempel's resolution of his paradox was to distinguish confirmation relative to *empty* background *vs* confirmation relative to *background containing antecedent knowledge of the properties of the objects to be observed.* Specifically, if we're testing H, Hempel would have said that we won't get *any* support (one way or the other) *if we turn over a card that we already know cannot be a counterexample to* H. As such, Hempel's approach to the paradox will (seemingly) *not help at all* in "explaining" why subjects might (or "should") turn over card c. [I think Nickerson just misinterprets Hempel's resolution at this point. Humberstone seems not to, but he doesn't seem to fully grasp all of its implications either.] Presumably, then, Hempel would have said that any strategy that turns over *only* cards already known not to be a counterexamples to H would be *confirmationally useless.* That is, it would provide *zero* degree of confirmation, as Maher's *early* Carnapian analysis also implies. Note: Maher's *later* Carnpian analysis does *not* imply this! Indeed, Maher's later systems entail that $Ra \,\&\, Ba$ *must disconfirm* H, relative to Ba. As a result, both the early and the later Carnapian systems are *incompatible* with Nickerson's Bayesian analysis (see below), which presupposes that $Ra\&Ba$ *confirms* H, relative to Ba. And, presumably, Hempel (and the early Carnap) would have said that any strategy that turns over some irrelevant cards and some relevant cards is *no better than* a strategy that turns over only the relevant cards in that set. So, for a Hempelian, S_1 would be no more useful than S_2, and S_3 would be no more useful than S_4. If we let $u(S)$ be the "Hempelian usefulness" of a strategy S, then I presume we have the following ordering. $u(S_1) = u(S_2) < u(S_3) = u(S_4)$. This is *bad* news for our subjects, *not good* news! We want a "confirmational usefulness" ordering such that $u(S_1) > u(S_3)$, since that's the empirical choice ordering. As such, *Hempelian* (or early Canrpian) intuitions about confirmation are *not* going to be of much use to us in "explaining" (in any normatively compelling way) the patterns of response here. Moreover, later Carnpian systems won't undergird Nickerson's Bayesian analysis either. To my mind, this indicates a rather serious disanalogy between Hempel's raven paradox (as it was originally conceived) and the Wason selection task (as Humberstonian-Bayesians like Nickerson reconstruct it). There is another (more minor) disanalogy between the original Wason task and Hempel's paradox. Recall that Hempel's paradox must involve $\sim Ra \,\&\, \sim Ba$, not some stronger claim like $Wa \,\&\, Sa$. This is my point about *monotonicity* creeping in, above and beyond the assumptions required to generate the original paradox. In Wason's original task, we have stronger properties than $\sim B$ and $\sim R$, which (by analogy) also presupposes monotonicity. Humberstone and Nickerson fix this by weakening the properties involved in cards b and d.

While Nickerson's Humberstonian-Bayesian analysis of Wason is disanalogous to the original raven paradox, so are most of the modern Bayesian treatments themselves (as I explained a few weeks back). Modern Bayesians make all kinds of non-Hempelian assumptions in their resolutions of the paradox. What Nickerson does is simply adapt these modern (quantitative) Bayesian approaches to the Humberstonian-Wason task. So, putting disanalogies to one side now, if we broaden the domain of discourse (this is the case Nickerson's paper is all about), then it is not clear that there is an *a priori* determinable uniquely "most useful" strategy, since verification is no longer possible (and, here, I assume that Hempel's intuitions are *not* knowably correct, *a priori*). This does open the door for *multiple* "useful" strategies. But, *exactly* how will this help here? There are various Bayesian analyses of the raven paradox that involve (quantitative) "two stage sampling", in which the observer *does* know properties of the objects before they observe them (see Earman's discussion online, for instance). Nickerson adapts these sorts of Bayesian approaches to modeling "relative usefulness" in the sorts of examples traditional Bayesians use to "soften the impact" of the paradox. He defines an interesting notion called *expected confirmational impact*, which is simply the expected degree of confirmation (he uses the difference measure, but in this context it doesn't matter much) of a single-card-turning strategy (this is his measure of "confirmational usefulness"). And, he gives examples (these are just like many of the standard quantitative Bayesian models one finds in the two-stage sampling literature) in which the ordering of the "usefulness" of single-card-turning-strategies is: $u(a) > u(c) > u(d)$, which seems to match what is observed.[1] This is a very nice analysis, which makes it possible for a Bayesian of the standard ilk to provide a clear sense in which it would be "reasonable" to prefer turning card c over turning card d, if your goal is to *maximize expected confirmational power* of the resulting evidence (relative to a background $\mathbb{K}$ that includes information about one side of all four shown cards, as in the example). This is very neat, but one wonders if the approach could be generalized, without making the standard sorts of independence assumptions, *etc.*, that Bayesians tend to make (which also force them to say various other things they may not want to say, in a comparative story). I haven't had time to work on that, but my *Mathematica* notebook on Nickerson (now posted) is a useful tool for beginning to think about modifications of Nickerson's "standard Bayesian" models. I do want to mention one other worry I have about his story.

In the case of the raven paradox, the assumption that $\Pr(\sim Ba \mid \mathbb{K}) \gg \Pr(Ra \mid \mathbb{K})$ seemed very plausible. In Nickerson's setup, he assumes that $\Pr(\sim Ta \mid \mathbb{K}) \gg \Pr(Ta \mid \mathbb{K}) > \Pr(Da \mid \mathbb{K})$. But, one wonders whether this assumption is as plausible in the Wason test. Why would subjects assume that there are more "3" cards in the "deck" than "D" cards (and far more non-3 cards than either of those)? In the ravens case, it is plausible that there are more black things than ravens (and that there are far more non-black things than either of those). But, I don't see the plausibility of the analogous assumption(s) in the Wason test. I think Nickerson's idea is that "ecologically" people tend to encounter universal generalizations in which these sorts of distributional assumptions hold. And, so, when faced with a novel one, they tend to revert back to these "default presuppositions" about the frequencies, etc. This may be plausible, but it's not clear to me.

2.2 Other Similar Explanations

Nickerson's explanation relies heavily on various assumptions about the sizes of the sets involved in the universal generalizations at hand. There has been some research suggesting that set-size does matter (and in the ways a standard Bayesian analysis would predict). But, more research on this needs to be done. Another reason to think set-size matters in these ways is the consideration that (as Nickerson puts it):

> If people interpret the selection task as "a laboratory model of the real-world task" . . . their failure to select $\sim Q$ is an analogue of the reasonable decision not to look for a nonblack raven by searching the universe of nonblack things. The assumption that there are many more nonblack things than ravens in the world

[1] He makes an error in his calculation of $u(c)$ (see my *Mathematica* notebook), but correcting it only bolsters his analysis. There is one possible worry here, which is that the strategies in Wason's test do not merely involve turning over single cards. And, it requires a more complex Nickerson-style Bayesian model to handle the full strategies S_i above. That could certainly change things. But, this is certainly a *good start* toward a more complete explanatory model, along standard contemporary Bayesian lines.

is critical to this line of reasoning. If it were the case that ravens outnumbered nonblack things, then the more efficient strategy would be to search the set of nonblack things, checking each to see if it is a raven.

Oaksford and Chater (1994) give a similar analysis to Nickerson's using *expected gain in information* about the hypothesis in question (now posted on website). This is *very* similar to Nickerson's account, since expected gain in Shannon information is ordinally equivalent to expected degree of confirmation (using measures d, r, or l) in these sorts of cases. What's "new" in Nickerson (*modulo* the Humberstone paper, which does no Bayesian analysis) is the connection with Hempel's paradox and Bayesian confirmation theory. Moreover, the detailed assumptions Oaksford and Chater make about the probability distributions involved are different in some interesting ways from Nickerson's models.

Cosmides and Tooby performed variations on this task that involve rules with some sort of "normative force". For instance, "All people who drink alcohol are over 21". If people are given this task (with four sample people who they are told drink alcohol, drink diet coke, are 23 years old, are 19 years old), they are much better at picking the correct pair to "test". But, in this sort of example, it is not clear whether set-size is also playing an important role, since that would also predict that people do better in this case (assuming that the domain of discourse is all people, of course — since there are more alcohol drinkers, I bet, than there are people under 21). This may also be because such "normative conditionals" make it clearer to people that the conditional is asymmetric and that the antecedent and consequent are what they're intended to be.

Moreover, people tend to look for positive instances (or confirming instances), rather than disconfirming or negative instances. And, people tend not to be very good at *modus tollens* reasoning, even if they are good at *modus ponens* reasoning. Again, that could have to do with its "ecological utility".

2.3 Nickerson's Models

Nickerson gives three different probability models, each of which is designed to compute the "expected degree of confirmation" of various one-card-turning strategies — assuming the standard contemporary quantitative Bayesian models for the raven paradox. Nickerson talks about a deck of cards on which raven/non-raven is written on one side and black/non-black is written on the other side. Then, he draws four cards from the deck and puts them on the table in front of you: r, $\sim r$, b, $\sim b$ (he uses lower-case letters for the visible side of the cards, and capital letters for the invisible sides). Now, the question he asks is: given the standard Bayesian models (Pr) for Hempel's paradox, what are the expected degrees of confirmation of the four possible one-card strategies: $u(a)$, $u(b)$, $u(c)$, $u(d)$? These quantities come out, as follows (here, I assume that $\mathfrak{c} = d$, and that it has been conditionalized on the subjects' actual background knowledge $\mathbb{K}$ about the raven set-up, as modern Bayesians assume it to be — Nickerson does a good job laying that out, but he doesn't fully specify his models properly, which leads to a more complicated story than he needs):

- $u(a) \overset{\text{def}}{=} \Pr(B \mid r) \cdot \mathfrak{c}(H, B \mid r) + \Pr(\sim B \mid r) \cdot \mathfrak{c}(H, \sim B \mid r) \approx 0.25.$
- $u(b) \overset{\text{def}}{=} \Pr(B \mid \sim r) \cdot \mathfrak{c}(H, B \mid \sim r) + \Pr(\sim B \mid \sim r) \cdot \mathfrak{c}(H, \sim B \mid \sim r) \approx 0.013.$
- $u(c) \overset{\text{def}}{=} \Pr(R \mid b) \cdot \mathfrak{c}(H, R \mid b) + \Pr(\sim R \mid b) \cdot \mathfrak{c}(H, \sim R \mid b) \approx 0.125.$
- $u(d) \overset{\text{def}}{=} \Pr(R \mid \sim b) \cdot \mathfrak{c}(H, R \mid \sim b) + \Pr(\sim R \mid \sim b) \cdot \mathfrak{c}(H, \sim R \mid \sim b) \approx 0.013.$

Thus, on the standard Bayesian models of Hempel's paradox (recast in Wason-style), we get $u(a) > u(c) > u(b) \approx u(d)$, which is (basically) the ordering of strategies we observe (when we "mod-out" by the other cards involved in the full strategies chosen by subjects). This is certainly a really interesting analysis, that goes a long way toward explaining what might be going on here. But, it does presuppose that the distributional assumptions about the raven case also apply to the Wason task (which I guess he argues for on "ecological default" grounds). It would be nice to (1) generalize this to the full strategy case, and (2) to find weaker assumptions than the standard Bayesian ones which suffice to ensure the desired u-ordering we see in the full experiments (as opposed to the one-card-turning simplification of it). This would make for a really nice paper/project, as it would be illuminating for both "ornithiology strategies" as well as strategies in the Wason selection task (understood as involving a broad domain of discourse, with a similar structure to that of the raven paradox case). I plan to write about that in my book.