Class 16 - Free Will III
Nichols and Knobe, "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions"

I. The Descriptive and the Normative

The final paper that we will discuss on free will and determinism, by Nichols and Knobe, is sensitive to the problems we have been discussing about the relevance of people's intuitions to the correct theory of free will.
Nichols and Knobe are careful to circumscribe all of their results.
They do not draw conclusions for either metaphysics or ethics from the data about folk theories of free will.

> The fact that people sometimes have compatibilist intuitions does not itself give us reason to suppose that compatibilism is correct (119).

> [The affective competence model] says... that people's responses in these conditions reflect a successful implementation of their own underlying system for making responsibility judgments. This claim then leaves it entirely open whether the criteria used in that underlying system are themselves correct or incorrect (120).

Their sensitivity to the question of how and whether folk theories of free will relate to philosophical theories of free will informs the entire paper.
They are also sensitive to some of the worries we discussed about the structure of some of the studies.

> We think it's a mistake to use technical terminology for these sorts of experiments, and we therefore tried to present the issue in more accessible language (110).

Nahmias et al. similarly avoided mentioning determinism, and only used 'free will' in one of the questions; they spend very little time on the responses to questions about free will in their paper.
In addition, Nichols and Knobe adjusted their scenarios so that the descriptions would be shorter, and clearer.
On the other hand, Nichols and Knobe did not verify whether subjects fully understood the ramifications of determinism, even when described non-technically, as Noah described in class.

II. Contextualism

Nichols and Knobe start by establishing that folk theories are contextualist, as Nahmias et al. considered.
In particular, they show that the emotional, or affective, aspect of a scenario changes people's ascriptions of moral responsibility.
They compare a concrete case of a man who kills his family (high affect) with an abstract case in which a deterministic universe was merely described (low affect).
Ascriptions of moral responsibility would indicate compatibilism, since the universe they describe is deterministic.
Failures to ascribe moral responsibility would indicate incompatibilist intuitions.
Subjects tended to be compatibilist (ascribing moral responsibility) in the high affect cases, and incompatibilist (denying ascriptions of responsibility) in the low affect cases.

Two factors distinguish the cases with higher ascriptions of moral responsibility.
First, they are concrete.
Second, they are high affect.
One might wonder whether the differences in ascriptions were due to the concreteness or to the affect.
Thus, in a later case, Nichols and Knobe asked whether subjects would be more willing to ascribe moral responsibility to a rapist than to a tax cheat, when the universe is deterministic.
In this survey, they controlled for concreteness, giving concrete stories about both the tax cheat (low affect) and the rapist (high affect).
Again, they found that subjects were far more willing to ascribe responsibility in situations of high affect.
Our ascriptions seem to depend on whether we are upset about an action.

> It seems that certain psychological processes tend to generate compatibilist intuitions, while others tend to generate incompatibilist intuitions (119).

Nichols and Knobe's contextualist conclusion continues to hold in the last experiment they report.
In the last survey, they demonstrated to subjects the incompatibility of the results from prior experiments.
They showed subjects that other subjects had given incompatibilist responses in abstract cases and compatibilist responses in concrete cases.
They emphasized to the new subjects that such responses contradict one another.
To resolve a contradiction, one must give up one of the contradictory claims.
Nichols and Knobe asked those new subjects whether, on reflection, they preferred compatibilism or incompatibilism.
Again, the results supported contextualism: to achieve consistency, some subjects chose to give up the compatibilist responses while others gave up the incompatibilist responses.
The folk just don't speak univocally about free will.

III. Accounting for Contextualism

Nichols and Knobe consider three classes of explanations of subjects' contextualist intuitions about free will.

1. Performance error
2. Affective competence
3. Concrete competence

In performance error, the difference in ascriptions is explained by attributing a distracting effect to the affective component of the scenario.
If the context-sensitivity of people's ascriptions is explained by performance error, then people's underlying intuitions could be incompatibilist.
Their emotional reactions to some scenarios distracts them into ascribing moral responsibility, in contrast to their more fundamental incompatibilist theory/intuition.
If compatibilist responses are a result of performance error, we should ignore results derived from scenarios with high affective content.
Ignoring emotional responses in favor of impartial, objective analysis is not unprecedented.
For example, we ordinarily want our jurists to be impartial, rather than emotional, in order to achieve objectivity.
But, there are other possible explanations of the contextualism that Nichols and Knobe discover.

An affective competence explanation, in contrast to performance error, takes the affective component as a trigger for our true intuitions.
On affective competence, subjects only truly express their views in high affect cases.
In low affect cases, they don't really see the question.
The affective competence model relies on some results in psychology that show that subjects with damaged emotional centers have difficulty grasping and using moral concepts.
Both performance error and affective competence models agree that one or the other response should be taken as reflecting our true intuitions.
They disagree over which responses are veridical.

A third attempt, the concrete competence model, explains the difference in ascriptions by aligning our intuitions with our responses to concrete cases.
In abstract situations, concrete competence supposes, subjects do not produce a legitimate response.
Concrete competence is clearly contravened by the comparison between the tax cheat and the rapist.
That survey held concreteness constant and still found differences in ascription correlated with affect.
But, concrete competence would be supported by a modular theory of mind vis-a-viz moral ascriptions.
If we have a specific module for making moral ascriptions, it might only be triggered by concrete cases; abstract cases might not set our real moral theory in motion.
Despite the evidence against concrete competence from the rapist/tax cheat example, it might be part of an explanation, part of a hybrid theory of our contextualist theories of free will.

Still, if we eliminate concrete competence, we are left to choose between performance error and affective competence to explain the Nichols and Knobe data.
Note that, in the 2x2 study (rapist/tax cheat x determinism/indeterminism), ascriptions of responsibility to the tax cheat dropped from 89%, in the indeterministic universe, to 23%, in the deterministic universe, a precipitous decline (p 117).
In contrast, ascriptions of responsibility to the rapist dropped from 95%, in the indeterministic universe, only to 64%, in the deterministic universe.
Nichols and Knobe argue that this difference in decline favors the performance error explanation.
Affective competence does not explain why ascriptions of responsibility drop so far for the tax cheat.
Thus, performance error seems most plausible.

Nichols and Knobe are again careful not to make sweeping conclusions on the basis of their limited results.

> Although our experiment provides some reason to favor the performance error account of the compatibilist responses we found, it seems clear that deciding between the affective performance error and the affective competence models of compatibilist responses is not the sort of issue that will be resolved by a single crucial experiment. What we really need here is a deeper understanding of the role that affect plays in moral cognition more generally (118).

IV. The Return of the Normative/Descriptive Problem

We have been wondering whether any appeals to folk theories of free will should have any ramifications for our reflective theories of free will.
Nahmias et al. argued that such results are important because our theories of moral responsibility are, in some way, tied to ordinary beliefs about free will.

> There is a reason why philosophers appeal to ordinary intuitions and common sense when they debate about free will: they are interested in developing a theory of freedom that is relevant to our ordinary beliefs about moral responsibility (82).

Nichols and Knobe seem sympathetic.

> The experimental results do not serve merely to give us insight into the causal origins of certain philosophical positions; they also help us to evaluate some of the arguments that have been put forward in support of those positions. After all, many of these arguments rely on explicit appeals to intuition. If we find that different intuitions are produced by different psychological mechanisms, we might conclude that some of these intuitions should be given more weight than others. What we need to know now is which intuitions to take seriously and which to dismiss as produces of mechanisms that are only leading us astray (119).

Given our framework of seeking reflective equilibrium, we have been willing to accept all sorts of intuitions, expecting that our mature theories of free will will accommodate only the proper intuitions.
We have been concerned to privilege the philosopher's intuitions, only because they are more likely to be in line with a coherent and reflective theory about free will.
Nichols and Knobe are working in the other direction, looking for reasons to privilege certain intuitions based on their causal origins, rather than on their coherence with a mature theory.
There is no *a priori* reason why one could not work in both directions at once, looking for principled reasons to approve of certain intuitions, and looking for a theory which preserves mainly the best intuitions.

Nichols and Knobe seem to think that there is reason to worry about the possibility of reaching reflective equilibrium, either for philosophers or for the folk.
At reflective equilibrium, they imply, we would have consensus on our intuitions and our best theories of responsibility and free will.
But subjects could not reach consensus even when asked to account for the discrepant intuitions underlying the weird Nichols and Knobe data.
Some subjects preferred compatibilism.
Some subjects preferred incompatibilism.
It is difficult to know what to make of this last study.
The n (=19) is low, so the results are not robust.
If we are to take the contextualism that Nichols and Knobe find as undermining Stich's so-called IDR strategy, we have to believe that such discrepant intuitions are robust.
The question is whether to take these subjects as in reflective equilibrium.
If they are, then the variability of intuitions is problematic.
But, one might argue, perhaps following Cohen, that evidence of discrepant or competing intuitions is precisely evidence against those subjects being in reflective equilibrium.