

Class 18 - Against Intuitions  
Gendler, "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium"

I. The Abstract and the Concrete

Gendler's paper centers on a phenomenon regarding our intuitions. Our responses to abstract scenarios seem to differ from those of concrete but relevantly-similar scenarios. Gendler surveys a variety of illustrations of the point. She also asks about the relevance of this phenomenon to philosophical methods.

We saw a version of Gendler's claim in the Nichols and Knobe paper. Abstract questions about people's free will in a determined universe tended toward incompatibilism, but questions about moral responsibility in concrete cases tended toward compatibilism. More precisely, Nichols and Knobe showed that people's intuitions about free will varied with their associated affect. When they read a scenario which generated an affective reaction, they were more likely to ascribe moral responsibility. Subjects appeared to be compatibilists. But when they read a scenario which did not generate an emotional response, they were less likely to ascribe moral responsibility. Subjects appeared to be incompatibilists.

Nichols and Knobe present a variety of different models to account for the difference. The central question for deciding among the different models is whether the increased affect pushes subjects toward better responses or whether the emotional reaction distracts them.

In his work on intentionality, Knobe pursues a different approach. He claims that apparent inconsistencies in subjects' responses may be deflated if we re-conceptualize the questions we are asking. Applying Knobe's method to the question of free will, we can deny that being both compatibilist and incompatibilist is inconsistent if we take our responses about free will to be tracking questions about moral responsibility. But if subjects are really mainly interested in ascribing responsibility, then their responses, which appear to concern a metaphysical question, are really about moral attributions. The increase in affect thus would appropriately track our increase in moral condemnation. The cases which appear to be relevantly similar are not similar because the differences in the scenarios' ability to raise affect is now a relevantly dissimilar factor.

Gendler's paper raises a similar question about our different responses to abstract and concrete scenarios. We will return to our central question, about how to understand the importance of differences in responses, after reviewing a few of the cases she cites.

## II. Historical Cases

In Hume's case of Themistocles, the Athenians reject a proposal to gain advantage over their rivals. The proposal is to set fire to the fleets of their competitor states.

But, the Athenians only hear of the proposal abstractly, described merely as an unjust act.

Hume's allegation is that if they were to have heard of it concretely, if they were to have heard the details of the plan and of the likely beneficial results, they would have acted.

Hume's case is hypothetical and contains no real data to support his conclusion.

But, Gendler takes Hume's interpretation as precedental.

If true, it supports her claim that concrete considerations have different affects than abstract ones.

By presenting content in a suitably concrete or abstract way, thought experiments recruit representational schemas that were otherwise inactive, thereby evoking responses that may run counter to those evoked by alternative presentations of relevantly similar content (69).

In the Themistocles case, the Athenians are supposed to miss an opportunity because they do not hear it expressed concretely.

In the David and Bathsheba case, David is convinced of the wrongness of his action by hearing a concrete case where appeals to abstract principles do not convince him.

David exhibits first-person bias in his actions, taking another man's wife and sending the cuckold off to die in battle.

When David is told a story of someone else acting in ways that are, in form, roughly equivalent to his own, he is able to recognize the wrongness of his action.

We are supposed to infer that abstract principles would not have convinced David.

Instead, he is convinced by the concreteness of the parallel case.

God recognizes the deep human tendency toward first-person exceptionalism, and seeks a way to speak to David that will circumvent this tendency (81).

The David case, like the Themistocles case, involves hypothetical explanations rather than data.

But, Gendler has data too.

## III. Logic and Mathematics

Gendler cites Thorndike's early-twentieth-century observation that people's ability to make mathematical inferences is oddly affected by adding subscripts to variables or changing variables to constants.

Changes in concrete particulars alter our abilities to do mathematics.

In the Wason test, subjects are presented with cards with letters on one side and numbers on the other.

They are asked which cards are relevant to testing a general rule like, "If there is an 'A' on one side, there is a '3' on the other."

Subjects often choose to look at a card showing a '3', which is not relevant to the rule.

Subjects reliably miss the fact that they should look at cards showing other numbers; if they have 'A's on their reverse sides, then they will be counterexamples to the rule.

That's the abstract case.

In the concrete case, in contrast, subjects tend to get the test right.

The concrete case invokes a rule such as, "If a person is drinking beer, then the person must be at least 21 years old."

In this case, subjects are not distracted by cards that say '21'; they reliably ignore them.

And, they do turn over cards that say, for example, '16', since they understand that if they say 'beer' on the reverse, they will be counterexamples to the rule.

The concrete nature of the case thus elicits better responses from subjects.

Gendler rightly points out that there has been a lot of research in the 45 years since Wason's original work.

It remains unclear exactly what accounts for subjects' distraction, despite a wealth of energy poured into research on the task, and a variety of explanations offered.

None of these accounts has been universally accepted, and it seems likely that the full story will turn out to be quite complicated (72, fn 5).

Still, the difference between the abstract version and the concrete version appears to be salient.

#### IV. Concrete Cases in Ethics

We talked earlier in the semester about how Rawls's *A Theory of Justice* returned first-order ethics to a position of respectability by defending the method of seeking reflective equilibrium.

Gendler credits Rawls with making moral theorizing concrete.

She claims that by invoking the thought experiment of going behind the veil of ignorance, Rawls derives his principles of justice from concrete situations.

We are more willing to engage the project just because it is concrete.

In Rawls's case, the scenario presented serves as a device of representation: It exhibits in vivid fashion the notion that Rawls takes to lie at the core of the concept of justice (85).

Similarly, Gendler cites Elizabeth Anderson's related mountain-climbing thought experiment as effective because of its concrete components.

Rawls inquires about the principles of justice we should choose for our society: principles for distributions of rights and responsibilities and social goods.

Anderson asks more specifically about principles of resource distribution.

She depicts a group of people, climbing a mountain, connected by ropes, and asks what principles for progressing, for balancing safety with aspirations, they would adopt.

In Anderson's case, the scenario presented serves as a tool for clarification: It encourages the reader to think through questions of risk and reward that Anderson takes to be common between her climbing game and the social structures governing resource distribution (ibid).

In Anderson's case, the concrete case of the mountain climbers is supposed to elicit our views about how much risk we are willing to adopt in order to gain advantages we seek.

Gendler sees both Rawls and Anderson as providing concrete instances of abstract general principles.

For both the Rawls case and the Anderson case, we could approach the questions more generally, more

abstractly.

My first response to the Anderson case, actually, was that it was a way of making Rawls's more-abstract scenario more concrete.

But in both cases, by concretizing the central questions, we are able to elicit intuitions that might otherwise not get expressed.

## V. Whither Intuitions

The question we must examine, in response to Gendler's observations, is whether the intuitions that concrete cases elicit are more accurate than our responses to abstract questions or whether the concrete nature of the cases is distracting us.

Are we finding performance errors or concrete competence?

Gendler mainly just makes a humble suggestion that there is a difference.

Decades of research in cognitive psychology have demonstrated that when content is presented in a suitably concrete or abstract way, this may result in the activation or fortification of a representational schema that was otherwise inactive or subordinate; the result of this may be to evoke responses that run counter to those evoked by alternative presentations of relevantly similar content (75).

As philosophers, though, we are interested in constructing true theories about the world.

We want to know which logic is correct and which principles of justice we should adopt.

If our responses to cases vary with their abstractness, we must get a grip on whether we should rely on our intuitions in abstract cases or in concrete ones.

The question of which intuitions we should find most reliable does not end with observations about concrete and abstract formulations.

Not all odd differences in intuitive responses are due to switching between abstract and concrete cases.

In the famous [trolley problem](#), due originally to Philippa Foot, but made popular by Judith Jarvis Thomson, people are more willing to throw a switch to redirect a trolley toward an innocent person than to push a fat man over the railing to stop the trolley.

The hands-on nature of the fat-man case turns people away from an act which seems to have the same moral worth as an act they are ordinarily willing to perform.

It appears that whereas the original trolley case produces increased neural activity in "higher cognitive" regions of the brain, cases such as fat man (where the imagined action is "up close and personal") produce increased neural activity in "emotional/social" regions (77).

Some philosophers react to the differences in intuitions between the fat man case and the switch-throwing case by wondering if there is, contrary to appearances, really a moral difference in the two cases.

We can understand Knobe's re-framing of the purpose of ascribing intentions to others as taking this route.

Knobe hypothesizes that our different ascriptions of intentionality in the help and harm cases are justifiable, despite appearances, if the real goal of such ascriptions is the assignment of blame.

The cases appear to be similar.

But, since one is blameworthy and the other is not, they actually are not relevantly similar.

Rejecting Knobe's approach, Gendler cites a variation in the trolley case which provides two scenarios which replace the fat man with a named individual.

In one scenario, subjects are asked whether they will push Chip Ellsworth III over the railing to save the Harlem Jazz Orchestra.

In the other scenario, subjects are asked whether they will push Tyrone Peyton over the railing to save the New York Philharmonic.

Whether or not there is a moral difference between the original trolley case and the fat man case, it seems clear that there is no moral difference between sacrificing Tyrone and sacrificing Chip. But if our only basis for thinking that there is a moral difference between fat man and original trolley is that subjects tend to respond differently to them, we should be disturbed to discover that parallel differences can be evoked by what seem clearly to be morally irrelevant differences (78).

In other words, variations in intuitions may not show us anything at all about morality. They may just reveal that we have a variety of psychological mechanisms which are triggered differently and which should not have any bearing on our best theories.

Not all intuitions are highly variable.

In cases of sense experience, our intuitions are stubborn.

Check out this talk from [Dan Ariely](#) about the persistence of intuitions.

See 2:25 - 5:00, especially, but the whole thing is great; he does the organ-donation and *Economist* cases we discussed earlier in the term.

Ariely points out that in cases of visual illusions, even when we understand the illusion, we often can not help but see the illusion persisting.

Similarly, in Gendler's jellybeans case, subjects chose worse odds despite their understanding of the odds.

Subjects choose to go with what they knew to be faulty intuitions, despite the presence of incentives (money) for choosing correctly.

Overall, more than eighty percent of subjects made at least one nonoptimal choice in the five trials each faced. When asked about their selections, "subjects reported that although they knew the probabilities were against them, they felt they had a better chance when there were more red beans... They made statements such as, 'I picked the ones with more red jelly beans because it looked like there were more ways to get a winner, even though I knew there were also more whites, and that the percents were against me'" (Denes-Raj and Epstein 1994, 819, 823) (73).

We might take the type of data we saw in Idiotfest as of this sort.

We have evidence of the variability of intuition and evidence of its constancy.

And we have evidence of the centrality of intuitions in philosophical reasoning.

Gendler argues for a disheartening conclusion.

Exactly because they recruit heretofore uninvolved processing mechanisms, thought experiments can be expected to produce responses to the target material that remain in disequilibrium with responses to the same material under alternative presentations, so that a true sense of cognitive equilibrium will, in many cases, prove elusive (69).

If we are seeking reflective equilibrium and our responses to what seem to be relevantly similar cases are contradictory, then we seem to be in a quandary.

Something has to give, but it is not clear which intuitions, or which theory, we should prefer.

As we see in the logic and mathematics cases, sometimes appeals to concrete cases help, but sometimes they hinder.

Subjects' tendency to reason validly is typically improved when materials are presented with familiar content, though there are also cases where familiar content may interfere with their ability to identify valid structures (71).

We know that in the process of seeking reflective equilibrium that we may have to cede some of our intuitions.

But we tend to reject philosophical theories when they conflict with our most resilient intuitions.

Gendler reports on the variability of intuition in Gettier-style and Truetemp cases.

People are more likely to ascribe knowledge when primed with clear cases of non-knowledge.

People are more likely to withhold ascriptions of knowledge when primed with clear cases of knowledge.

Perhaps we have to re-think our approaches to such philosophical theories.

If Gendler's interpretation of the data she surveys is correct, we are going to have to expect that even our most heartfelt intuitions may have to be given up.

A philosophical theory may be correct even if we consistently and resiliently react to specific cases in ways that run counter to the theory's predictions. This introduction of an additional degree of freedom into the enterprise of philosophical explanation may introduce a feeling of vertigo (76).

A better understanding of the factors influencing our intuitions seems required before we can make any useful observations relying on our intuitions.