AFFECT AND RESPONSIBILITY
I. Nichols and Knobe’s study

- Kane (along with most philosophers) maintains that our ordinary intuitions favor incompatibilism, but a recent study by Nahmias et al casts doubt on this.

- However, Nichols and Knobe hypothesized that it made a difference whether or not the question was posed in terms that engage our emotions.

- Their study (see pp. 12f.) solicited responsibility judgments about cases described either
  - in terms of a *concrete* example of wrongdoing, or
  - in *abstract* theoretical terms, as in a philosophy class.
Initial findings

- Their results (see pp. 12-13) confirmed Nichols and Knobe’s expectation of a divide in our intuitions:
  - We have **compatibilist** intuitions in response to more concrete questions about particular transgressions.
  - We have **incompatibilist** intuitions in response to abstract questions about a deterministic universe.

- However, on the basis of this first experiment alone, Nichols and Knobe were unable to explain the divergence as a result of the influence of concrete expression on emotional affect [= the felt aspect of emotion – as opposed to accompanying thoughts, behavior tendencies, etc.].
Nichols and Knobe outline three main psychological models that (along with hybrid models) might explain their results:

1. **affective performance error**: Emotions distort our judgment, so that compatibilist responses don’t really indicate a compatibilist view, but just an error in responsibility attribution. Our intuitive theory is **incompatibilist**.

2. **affective competence**: Emotions are part of the basis for moral judgment, and hence for judgments of responsibility, as manifested in **compatibilist** responses.

3. **cognitive competence**: Emotions play no role, but the purely cognitive “modular” process that generates responsibility judgments applies only to concrete situations. So the findings would again mean that our intuitive view is **compatibilist**.
Further testing

- A further experiment found that responsibility attributions were much higher when a concrete case was described in a way that evoked high vs. low affect (with rape as the crime vs. cheating on taxes).

- The results (summed up in the matrix on p. 23) showed that affect, not just concreteness, was the relevant factor, contrary to the cognitive competence model.
  - Nichols and Knobe take these results to fit best with the affective performance error model, on which we’re intuitively incompatibilists, but are misled by our emotions. But they still take the evidence for the model to be inconclusive.
  - Let’s examine the three models in reverse order to understand their conclusions.
Model 3: Cognitive competence

Emotions play no role, but the purely cognitive “modular” process that generates responsibility judgments applies only to concrete situations.

- This model seems to be disconfirmed by Nichols and Knobe’s second experiment.

- However, the model has evolutionary plausibility, so they don’t dismiss it completely. It would make out our competence as an innate adaptation to allow for rapid response.
Model 2: Affective competence

Emotions are part of the basis for moral judgment, and hence for correct judgments of responsibility.

- This model gets some support from studies of people with emotional deficits, who seem to have impaired moral judgment as a result.
- But how can it explain why responses to low-affect cases should depend on determinism? Could a normal competence fail to be triggered in such common circumstances?
Model 1: Affective performance error

Emotions distort judgment, so that compatibilist responses don’t really indicate a compatibilist view, but just an error in judgment.

- This does seem to explain Nichols and Knobe’s results. It would mean that the high-affect condition introduces bias that makes apply our intuitive theory of responsibility incorrectly.

- However, Nichols and Knobe’s favored hybrid model (see p. 20) still might apply. Maybe our competence depends on moderate affect.
Philosophical implications

- Nichols and Knobe stress that their experiments establish nothing about the truth of our intuitions, i.e. whether compatibilism or incompatibilism is true, as opposed to what we think is true.
  - For instance, taking an “affective performance error” to explain compatibilist responses might suggest that compatibilism is false, since belief in it is the result of an error.
  - But the error would be in applying our intuitive (incompatibilist) system of responsibility attribution, which might itself be mistaken.

- Their results have implications for the free will debate just insofar as they undermine some authors’ attempts to use compatibilist intuitions as evidence for their theories.
Interpersonal divergence

Nichols and Knobe report a final experiment indicating that there’s no one way in which people tend to resolve the conflict between compatibilist and incompatibilist intuitions.

- "Reflective equilibrium" is a term they take from moral philosopher John Rawls. It involves a back-and-forth adjustment of our theoretical intuitions and our intuitions on particular cases so as to eliminate inconsistency.

- About half of their subjects resolved the conflict in favor of compatibilism, whereas the other half favored incompatibilism.

- Their conclusion is that the experiment indicates a deep divide — and a need for further research on morality and emotion.
II. Watson on Strawson

- Strawson understands responsibility as constituted by [justified] reactive attitudes like resentment – without a need for independent judgments backing them up.
  - Watson refers to this as the expressive theory because it interprets reactive attitudes as expressing a moral demand.
  - Holding someone responsible thus has a communicative purpose – as distinct from the possibility of change in behavior stressed by consequentialist theories like utilitarianism.

- Watson’s aim in the essay is to see whether the account can be extended to answer problems raised by cases where a wrongdoer fails to make any sense of moral demands. Would Strawson’s view require adopting the “objective” attitude, in place of blame?
Evil as an exemption?

- Strawson distinguished between two types of “pleas,” i.e. reasons why someone should not be held responsible:

  1. **excuses** (where the agent didn’t really violate the moral demand in question – e.g., he didn’t act deliberately)

  2. **exemptions** (where the agent isn’t an appropriate object of a moral demand – e.g. he wouldn’t understand it, or already accepts it but “wasn’t himself” when he acted)

- Watson interprets these as indicating constraints on “moral address,” the communication of a moral demand.

- The problem is that the second category seems to include extreme cases of evil, where an agent apparently lacks the ability to comprehend moral demands.
The criminal acts of Robert Alton Harris

- Watson quotes at length from a newspaper account of Robert Alton Harris, a criminal on death row.

- Harris killed two youths in cold blood and casually ate their lunch, while musing on how much fun it would be to pose as police and inform their parents.

- At the time of Watson's article he was on death row, hated by the other inmates and still showing no remorse.

- In response to the description of Harris's crime and his current character, reactive attitudes of blame – "retributive" attitudes – seem to be justified by his ill will, in Strawson's terms, despite his imperviousness to moral considerations.
The relevance of Harris’s formative circumstances

- Strawson includes among exemptions “unfortunate formative circumstances,” i.e. bad luck in the conditions under which an agent’s character was formed.

- Watson goes on to quote Harris’s sister’s account of his gentle nature as a young child and his abused and unloved childhood.

- However, the fact that sympathy for Harris is *also* justified doesn’t seem to exempt him from responsibility.

- Instead, ambivalence is justified, i.e. feeling *both* sympathy and blame (vs. just one, or neither, or uncertainty).
“Moral Luck”

- What does undermine the justification for blaming Harris, according to Watson, is the realization that, if our own formative circumstances had been as bad as his, we’d have developed the same ill will.

- [The term “moral luck” refers to the dependence of responsibility on factors outside the agent’s control – not to be confused with the Luck Objection to libertarianism.]

  - One might ask why we don’t just grant that we’d be blameworthy too, if we’d had Harris’s bad luck.

  - But Watson’s point is that on Strawson’s “moral address” account, we need to be justified in blaming Harris as things stand. But we’re in no position to blame him if our morally superior position is just a matter of luck.]
A problem for Strawson’s compatibilism?

- Watson’s takes his discussion of moral luck to show that the Strawsonian account doesn’t really support compatibilism.
  - Strawson doesn’t allow any distinction between reactive attitudes and a judgment of responsibility, so both would be affected by moral luck.
  - We couldn’t say that it’s appropriate for us to hold someone responsible but not to feel ill will toward him ourselves.

- But determinism would seem to extend moral luck to all our actions, so it would undermine the justification for any reactive attitudes and hence rule out responsibility in Strawsonian terms.
Non-retributive responses

- Watson concludes that Strawson’s version of compatibilism won’t work, and he’s rejected libertarianism in earlier sections of the article. [Remember that his own account resembles Frankfurt’s compatibilism; see Kane, ch. 9.]

- But he recoils at the sense of personal isolation implied by Strawson’s “objective” attitude toward others we can’t hold responsible, which is what he thinks we’d be left with if we embraced hard incompatibilism.

- Instead, he favors detaching the judgment of responsibility from retributive attitudes, using the model of Gandhi and Martin Luther King, who apparently managed to hold others responsible while forgoing blame.