Reason, Emotion, and Moral Judgment

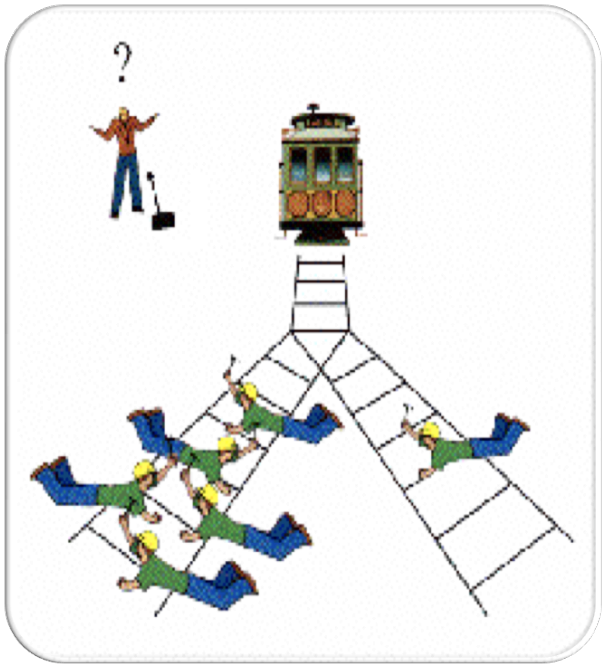
TFAP
2012/7/12
Yukihiro Nobuhara
The University of Tokyo

Abstract

Yukihiro Nobuhara, “Reason, Emotion, and Moral Judgment”

Neuroscientific study on moral cognition has demonstrated that emotion plays an important role in moral judgment. But what role does emotion play in moral judgment? It would not be surprising if emotion only distorts moral judgments. The interesting question is whether emotion at least sometimes promotes rational moral judgments. To address this question, I examine two models of moral cognition. One is Greene’s cognitive control model, in which emotion only distorts moral cognition. The other is Moll’s cortico-limbic integration model, in which emotion cooperates with reason to promote adequate moral cognition. I argue that Moll’s model is superior to Greene’s. Finally, although Moll’s model emphasizes the integration of reason and emotion in moral cognition, I argue that the phenomena of weakness of the moral will show the existence of a purely rational system.

Let's start with the famous trolley problem (Foot 1967)



From J. Greene's HP

The switch dilemma

- A runaway trolley is headed for five people who will be killed if it proceeds on its present course.

The only way to save them is to hit a switch that will turn the trolley onto an alternate set of tracks where it will kill one person.

- Ought you to hit a switch to turn the trolley in order to save five people at the expense of one?

- Most people say yes.

From J. Greene's HP

The footbridge dilemma

- As before, a trolley threatens to kill five people. But this time, you are standing next to a large stranger on a footbridge. You are light, and the stranger heavy; so your body cannot stop the trolley, but his body can.

The only way to save the five people is to push the stranger off the bridge onto the tracks below. He will die, but his body will stop the trolley.

- Ought you to save the five people by pushing the stranger to his death?

- Most people say no.

The interesting point:

- The switch dilemma and the footbridge dilemma have the same structure from a utilitarian point of view:

  to save five in the expense of one.

- But the answers are different.

- Why different?

One answer is Greene's:

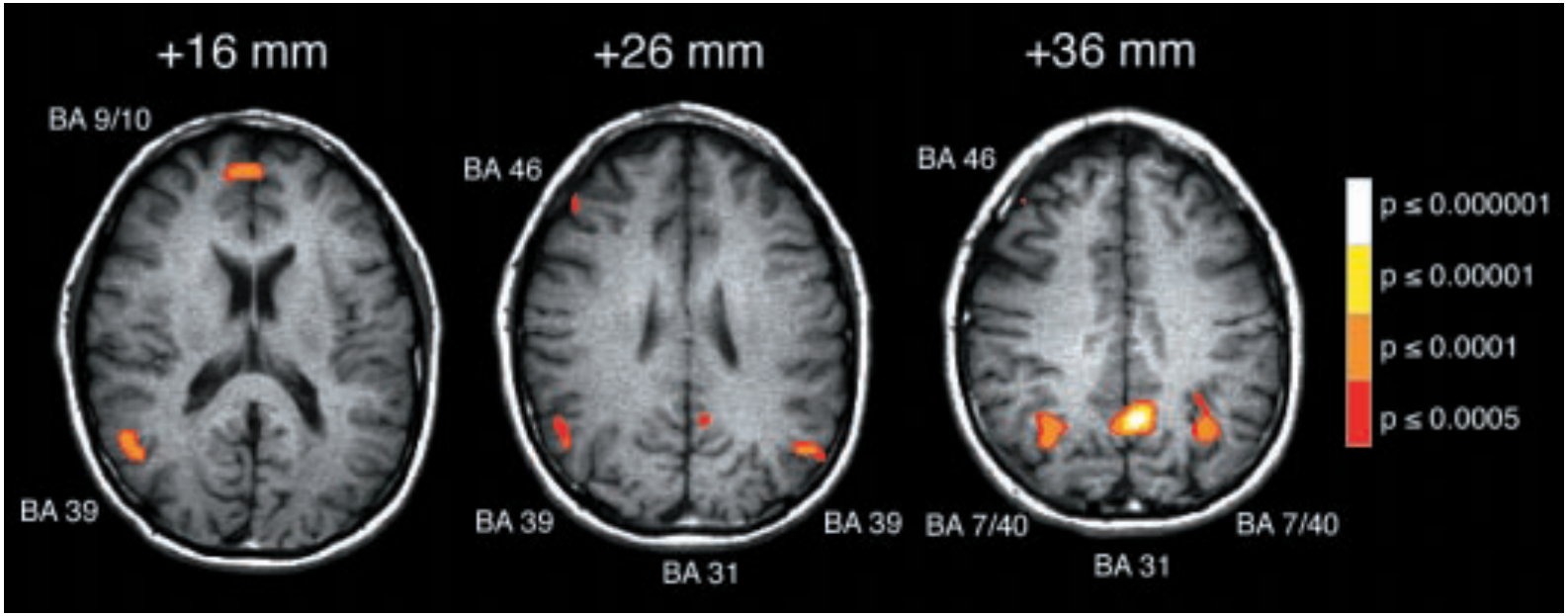
Greene et al. (2001) investigated neural activities of people who were tackling these two dilemmas.

The Results about neural activities:

Medial portions of Brodmann's Areas (BA) 9 and 10 (medial frontal gyrus), BA 31 (posterior cingulate gyrus), and BA 39 (angular gyrus, bilateral) were significantly more active in the footbridge condition than in the switch condition. All of these three areas are implicated in emotional processing.

Greene's interpretation:

The footbridge condition evokes stronger negative emotion than the switch condition because you have to directly push the stranger off the bridge.



If Greene's interpretation is right or at least on the right track, emotion probably plays some important role in moral judgment.

What role does emotion play in moral judgment?
-It would not be surprising if emotion only sometimes distorts moral judgment, since it is just common knowledge.
Ex. You judge that you need not help a child nearly drowning in the river in front of you, because you feel fear of yourself drowning if attempting to help the child.

-The interesting question is whether emotion sometimes enables right moral judgment.
That is, are there cases in which we cannot make a right moral judgment because we have no adequate emotions?
In other words, can we say that emotion sometimes plays a positive role in moral judgment?

Outline

1. Greene's cognitive control model: it assigns no positive role to emotion
2. Moll's cortico-limbic integration model: it assigns a positive role to emotion
3. Weakness of moral will and a purely rational system:

   I argue, based on the phenomena of weakness of moral will, that there is a purely rational system, contra Moll's model

Greene's Cognitive Control Model

Let's see in more detail Greene's interpretation of the experimental results on the trolley problem.

◆The distinction of personal-impersonal

First, Greene makes an important distinction: personal moral dilemmas and impersonal moral dilemmas.

This distinction is made by the three criteria:

1) the action must be likely to cause serious bodily harm.
2) this harm must befall a particular person or set of persons.
3) the harm must not result from the deflection of an existing threat onto a different party. (=the harm is not the by-product but the means to the end.)

The dilemmas meeting all the three criteria are personal, and others are impersonal.

- The switch dilemma is impersonal:
  + The harm results from the deflection of the existing threat (=the trolley's killing 5 people) on a different person at the different track.
  + So, the harm is the by-product of saving the five people.

- The footbridge dilemma is personal:
  + The harm results from directly pushing the stranger off the bridge.
  + so, the harm is the means for saving the five people.

◆ Greene's cognitive control model of moral judgment

Based on the distinction of personal and impersonal moral dilemma, Greene proposes the cognitive control model of moral judgment.

According to this model:

- Impersonal dilemmas are less emotionally engaging while personal ones are more emotionally engaging.
- So, Reason can control emotion in impersonal dilemmas, but not in personal dilemmas.
- So, the responses are different between these two types of dilemmas.

Let's apply this cognitive control model to the switch dilemma and the footbridge dilemma:

1) the footbridge condition

- the reason infers that it is better to push the stranger off the footbridge to save the five people,
- but a strong negative emotion toward this action occurs automatically,
- so the reason cannot control the emotion

→ "no" to pushing the stranger.

2) the switch condition

- the reason infers that it is better to hit the switch to save the five people,
- only weak or no negative emotion toward this action occurs,
- so the reason can, or need not, control the emotion

→ "yes" to hitting the switch.

* A note on terminology

In Greene's terminology

- the decision to save five at the expense of one: a utilitarian choice because it produces more total amount of happiness.

- the decision not to sacrifice one in order to save five: a deontological choice because it accords with the moral principle that one ought not to treat a person only as a means.

This terminology does not seem completely adequate,

but I will use it in what follows only for the sake of convenience.

◆ A problem with Greene's cognitive control model

In a nutshell, according to Greene's model,

- the utilitarian choice is supported by the reason,
- but the emotion against it occurs in personal moral dilemmas such as the footbridge dilemma,
- so the reason attempts to control the emotion in order to resolve the conflict,
- if the reason succeeds in the control, the utilitarian choice wins,
- but if not, the deontological choice wins.

In this way, in Greene's model,

the choice in personal dilemmas is the result of the conflict of reason vs. emotion and the cognitive control of emotion.

However, it is doubtful that in all personal dilemmas, such conflict and cognitive control really occur.

For example, in the infanticide dilemma, there would not occur any conflict.

* The infanticide dilemma
  + a teenage girl has a baby,
  + but she does not want the baby because the baby prevents her selfish life,
  + so she must decide whether to kill her unwanted newborn infant or not.

Question: Is it appropriate for the teenage mother to kill her child?

- The option for killing her own child would be immediately thrown away, because
  the negative emotional response towards it would be strong,
  and the cognitive reason for it would be weak,
  so the emotion would overwhelm the reason.
- So, there would be no conflict, so no control in this dilemma.

◆ Easy vs. difficult moral dilemmas

Greene himself realized this problem.

So, he divides personal dilemmas further into two subclasses: easy dilemmas and difficult dilemmas.

- In easy dilemmas, the choosing process takes a relatively short time and most people make the same choice.
- In difficult dilemmas, the choosing process takes a relatively long time and the choices split among people.

For example,

- The infanticide dilemma is easy.
- The crying baby dilemma is difficult.

*The crying baby dilemma:

- Enemy soldiers have taken over your village.

  They have orders to kill all remaining civilians.

- You and some of your townspeople have sought refuge in the cellar of a large house.

  Outside, you hear the voices of soldiers.

- Your baby begins to cry loudly.

  You cover his mouth to block the sound.

- If you remove your hand from his mouth, his crying will summon the attention of the soldiers, and they will kill you, your child, and the others hiding out in the cellar.

- To save yourself and the others, you must smother your child to death.

Question: Is it appropriate for you to smother your child in order to save yourself and the other townspeople?

◆Greene's prediction

After dividing personal dilemmas into easy and difficult ones, Greene makes the following predictions:

- In easy dilemmas,

the negative emotion toward the utilitarian choice is overwhelming

→ no conflict

→ no cognitive control

→ the deontological choice is immediately made

Ex. In the infanticide dilemma, it is immediately judged that the mother ought not to kill her newborn baby.

- In difficult dilemmas,

there emerges a conflict between the negative emotion toward the utilitarian choice and the rational reasoning for it

→ the cognitive control is made

→ the utilitarian choice is made if the control succeeds, and the deontological choice is made if it fails.

Ex. In the crying baby dilemma,

- you experience the conflict between the negative emotion and the rational reasoning about smothering your baby,
- and if the rational reasoning succeeds in controlling the negative emotion, you judge that you ought to smother your baby,
- and if it fails, you judge that you ought not to smother your baby.

◆ The experimental results

Greene made the experiments to test his predictions.

He investigated with fMRI the neural activities of the subjects when they were considering the easy and difficult dilemmas.

The results:

The dorsolateral prefrontal cortex (DLPFC) and the anterior cingulate cortex (ACC) are more activated in the difficult dilemmas than in the easy dilemmas.

(It is known in advance that ACC is involved in conflict and DLPFC in cognitive control.)

Greene interprets these results as showing that conflicts and cognitive controls arise in the difficult dilemmas while they do not in the easy ones.

Thus, Greene thinks that his predictions were supported by these results.

◆ The distinguishing features of Greene's cognitive control model

In Greene's cognitive control model,

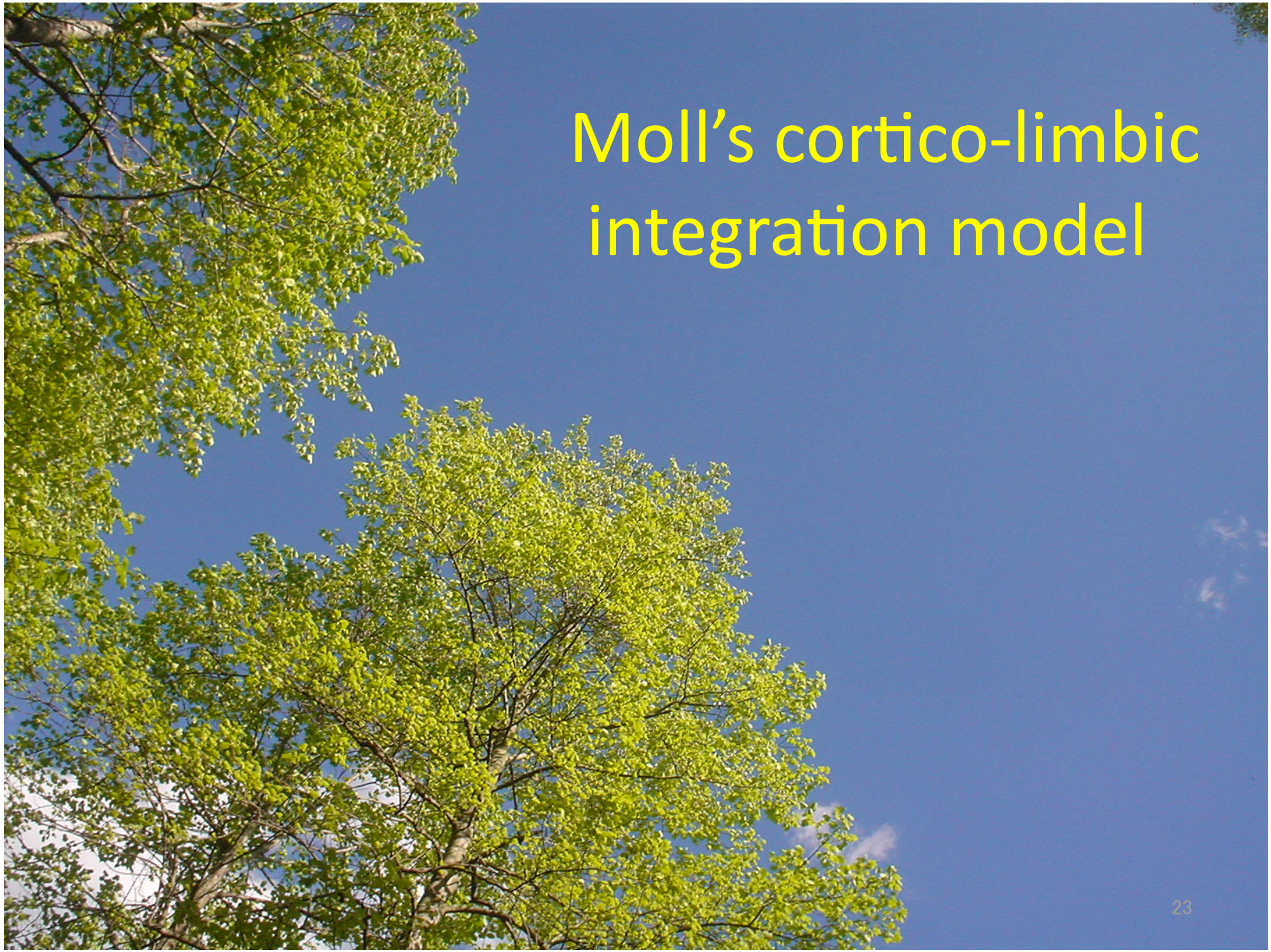
- Reason and emotion are in conflict.
- Rational moral judgment is made if reason succeeds in cognitive control of emotion.
- Therefore, emotion only distorts moral judgment; it does not play the role of promoting rational moral judgment.
- For example, in the case of the footbridge dilemma,

  the deontological decision not to push the stranger off the bridge is not rational, though most people make it,

  because, in this case, the reason fails to cognitively control the negative emotion toward pushing the stranger off the bridge.

◆ Is there any model which allows for a positive role for emotion in moral cognition?

Let's examine Moll's model of moral cognition for pursuing such a model.


Moll's cortico-limbic integration model

◆Moll's model

In Moll's model (Moll et al. 2005a, Moll et al. 2008),

- In moral cognition,

  the prefrontal cortex (involved in rational cognition) and

  the limbic system (involved in emotion)

  → shape an integrated system.

- In this integrated system,

  each option is evaluated in terms of both reason & emotion,

  and the most highly evaluated option is selected.

- For example, in the footbridge dilemma,

+ the option for pushing the stranger off the footbridge to save the five others is evaluated in terms of the following considerations,

its saving more people,

but causing the angst for being a murderer,

----------;

+ the option for not pushing the stranger is evaluated in terms of the following considerations,

its leading to avoidance of being a murderer,

but causing the guilt for doing nothing to save the five others,

----------.

+ If the option for not pushing the stranger is more highly evaluated, it is selected,

and if reverse, the option for pushing the stranger is selected.

- According to Moll,

+ the conflict between reason and emotion that Greene alleges is, in fact, nothing but the conflict between two rational-emotional alternatives which are very difficult to choose between,

+ and the control of emotion by reason is, in fact, nothing but a very bitter decision between them.

◆ The positive role of emotion in Moll's model

In the cortico-limbic integration model, not only reason but also emotion plays a positive role in moral cognition.

- In the footbridge dilemma, the negative emotion toward pushing the stranger off the footbridge is an adequate emotion.
- Such adequate emotions cooperate with reason to enable right moral cognition.
- Inadequate emotions such as a pleasure of killing a person would disturb right moral cognition, but adequate ones would promote it.
- This is just like reason; if reason functions adequately, it promotes right moral cognition, but if not, it disturbs it.

◆ The motivational role of emotion

Moreover, emotion, unlike reason, plays a role of motivation to actually perform a chosen course of action.

- Even if you choose to smother the crying baby to save yourself and the townspeople, you could not actually smother the baby unless you have any emotions supporting the choice.
- For example, if you have only the emotions opposing the choice such as the compassion for the baby, you cannot smother the baby.
- Thus, emotion not only sometimes contributes to right moral judgment, but also motivates us to actually perform the chosen course of action.

◆ Moll's criticism of Greene's model

Now let's consider which model of moral cognition is superior, Moll's cortico-limbic integration model or Greene's cognitive control model.

To consider this, it is helpful to examine Moll's interesting criticism of Greene's model (Moll et al. 2008).

This criticism is based on the results of Koenigs' experiments in the Ultimatum Game* with VMPFC patients* (Koenigs and Tranel 2007).

* The Ultimatum Game

- Two players are given one opportunity to split a sum of money (ex. $10).
- One player (the proposer) offers a portion of the money (ex. $4) to the second player (the responder) and keeps the rest ($6).
- The responder can either accept the offer or reject it;
  + if the responder accepts it, both players split the money as proposed (the responder 4$, the proposer 6$),
  + but if the responder rejects it, both players get nothing.
- Normal people usually accepts the almost fair offers (the half or a little bit less than the half) and rejects the unfair offers (much less than the half, ex. $1, $2).

* VMPFC patients

VMPFC (the ventromedial prefrontal cortex) is involved in emotional processing, so patients with damage to VMPFC are emotionally blunt.

◆ Koenigs' experiment

- Koenigs made the experiments in the Ultimatum Game with VMPFC patients.
- The result of the experiment: VMPFC patients tend to reject the offers of larger amounts of money (ex. 4$) than the normal subjects.

1) Now, rejecting any offer except the $0 offer causes a loss to the responder because the responder gets nothing in spite of being able to get some amount of money by accepting the offer

   (this is a one-off game, not a repeated one with the same person, so rejecting the offer does not lead to any later gain).

2) So the utilitarian choice is to accept any offer except the $0 .

3) If the responder rejects an offer in spite of a loss, it may be attributed to the resentment toward the unfair proposer;

   the responder is attempting to punish the proposer by making the proposer get nothing at the cost of the responder's own gain.

4) So rejecting an offer is a deontological choice.

- Thus the experimental result means that VMPFC patients has a stronger inclination to the deontological choice than the normal subjects.

◆ The apparently contradictory results

- In the Ultimatum Game, VMPFC patients has been demonstrated to have a stronger inclination to the deontological choice than the normal subjects.
- However, from other experiments in the difficult personal dilemmas , VMPFC patients are known to show a stronger inclination to the utilitarian choice,

  Ex. In the footbridge dilemma, they tend to choose to push the stranger off the footbridge.
- How can we explain these apparently contradictory results?

◆ Moll's explanation of the apparent inconsistency

Moll asserts that the explanation is impossible in Greene's model, but it is possible in his model. His explanation is:

- VMPFC is not involved in all kinds of emotion, but only in a restricted range of emotion.
  - + VMPFC may be critical for prosocial emotions such as attachment, compassion, and guilt;
  - + DLPFC (the dorsolateral prefrontal cortex) and lateral OFC (the orbitofrontal cortex) may be relevant for socially aversive emotions such as resentment, indignation and contempt. (Moll et al. 2005b)
- If so, VMPFC patients,
  - + in the difficult personal dilemmas, tend to make the utilitarian choice, i.e., to opt for saving many people by killing one person, because they lack the prosocial emotions due to the damage to VMPFC,
  - + in the Ultimatum Game, tend to make the deontological choice, i.e., to opt for rejecting even a little bit unfair offer because they retain the social aversive emotions with DLPFC and lateral OFC intact . (Moll et al. 2008)
- Thus, Moll asserts that his model can explain the apparently contradictory results with VMPFC patients.

◆ Moll's model is more plausible

- After all, which model is more plausible at present, Moll's cortico-limbic integration model or Greene's cognitive control model?
- On the one hand, it is not sure whether Greene's model can by no means explain the apparently contradictory results, and on the other hand, there are certainly a lot of problems with Moll's model, too.
- So, to have a definite answer to the question of which model is superior, or whether there is a third model better than both, we need a further advancement of neuroscientific study on moral cognition.
- But for the present, we can conclude that Moll's model is most convincing in terms of the current neuroscientific evidence and consistency.


Weakness of moral will
and a purely rational system

◆ The question: is there a purely rational system?

- In Moll's cortico-limbic integration model, reason and emotion are integrated in the cortico-limbic system, and each option is evaluated in terms of both reason and emotion;

  so there seems to be no room for a purely rational system.

- However, from our everyday experiences, it seems that there is a purely rational system distinct from the cortico-limbic integration system.

  The purely rational system does not include emotion;

  it consists only of reason, so it evaluates each option and makes a decision from a purely rational point of view.

◆ Weakness of moral will

- What suggests most clearly the existence of a purely rational system in our everyday experiences is the phenomena of weakness of moral will.
- For example, in the crying baby dilemma, the following case is possible:
  + I judge that I ought to smother the baby.
  + Yet I finally decide not to smother the baby, though I continue to hold the judgment that I ought to smother the baby.
  + That is, I retain the moral judgment and make a final decision contrary to it.
- This case can be called weakness of moral will because
  I make a decision contrary to my moral judgment,
  that is, I cannot shape my will according to my moral judgment.

◆The existence of a purely rational system

How does weakness of moral will suggest the existence of a purely rational system?

- Any decision or will has motivational force, so it is made in the cortico-limbic integration system, which has motivational force due to emotion.

  So, in the case of weakness of moral will, the decision not to smother the baby is made in the cortico-limbic integration system.

- But how about the judgment that I ought to smother the baby?

- If the judgment was made in the cortico-limbic integration system, it would be temporary, because the decision not to smother the baby is made finally in the system.

- So the judgment would not be retained when the decision not to smother the baby has been made.
- However, in the case of weakness of moral will, the judgment that I ought to smother the baby is retained when, and even after, the final decision not to smother the baby has been made.
- Therefore, the judgment cannot be made in the cortico-limbic integration system; it should be made elsewhere.
- So, it should be made in a purely rational system distinct from the integration system.
- Thus, there must be a purely rational system.

◆ How the purely rational system works

- In the purely rational system, both options, smothering/not smothering the baby, are evaluated cool-heartedly and rationally.
  + On the one hand, if I smother the baby,
    the baby will die but many other lives will be saved;
    to victimize the baby for saving many lives is to use the life of the baby as a means to an end, so it is morally very problematic;
    yet it is morally good that much more lives will be saved.
  + On the other hand, if I do not smother the baby,
    the enemy soldiers will find us, and all of us, including even the baby, will be killed;
    yet I do not commit the crime of killing the baby with my hand, as a means to an end.
- Thus, taking various considerations into account and weighing both options cool-heartedly, that is, without emotions, I finally judge that I ought to smother the baby.

- However, in the cortico-limbic integration system,
  + the options are evaluated in terms of emotion as well as reason;
  + so killing the baby with my hand may be evaluated as a more serious sin in the integration system where the thought of killing the baby causes a severe feeling of guilt.
  + so the decision not to smother the baby may be made in the cortico-limbic integration system.
- This is why the judgment made in the purely rational system may not accord with the decision made in the cortico-limbic integration system.

◆ The weakness of moral will and the executive power

- When the integration system makes a decision which disagrees with the judgment in the purely rational system, the decision in the integration system is put into action,

  Ex. When the purely rational system judges that the baby should be smothered and the integration system decides not to smother the baby, the action of not smothering the baby is actually carried out.

- In this sense, the integration system has the executive power, but why is this so?
- The reason is, as Moll asserts, that only emotion can provide motivation for action.
- Rational cognition concerns motivation only through being reflected in emotion.

  If the purely rational system is to affect behavior, judgments in this system must be reflected in emotions in the cortico-limbic integration system.

- If the reflection is not proportional, the integration system may make a decision which disagrees with the judgment in the rational system.

  This is how the weakness of moral will, or the weakness of will in general, occurs.

◆ Why a purely rational system exists?

- In this way, the integration system has the executive power, but nevertheless it seems that there exists a purely rational system distinct from the integration system;

  if so, why does it exist, or for what?

- The purely rational system can evaluate things in a more long-term and comprehensive perspective and therefore can take more universal and objective values into account, because it is not affected by emotions.
- Judgments in the purely rational system are not necessarily reflected proportionally in emotions in the integration system.

  This is because they are sometimes made in a too long-term or too comprehensive perspective.

- However, we can admit that the purely rational system has its own unique reason for existence, because

  it plays a distinct role from the cortico-limbic integration system,

  that is, evaluation in a more long-term and comprehensive perspective.

By the way,

- The purely rational system can be regarded as corresponding to rational cognition in Greene's cognitive control model.
- But unlike Greene's model, it does not control emotion;

  it does not confront an purely emotional system,

  rather it confronts the reason-emotion integration system.

Conclusion

In sum,

- Moll's cortico-limbic integration system is more plausible than Greene's cognitive control model.

- But contra Moll's model, there seems to be a purely rational system, given the phenomena of weakness of moral will.

Reference

- Foot, P. (1967) The Problem of Abortion and the Doctrine of the Double Effect. *Oxford Review*, 5, 5-15.
- Greene, J.D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., Cohen, J. D. (2001) An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science* 293, 14 SEPTEMBER, 2105-2108
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., and Cohen, J. D. (2004) The Neural Bases of Cognitive Conflict and Control in Moral Judgment. *Neuron* 44, October 14, 389-400.
- Haidt, J. (2001) The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review* ,108 (4), 814-834.
- Haidt, J. (2007) The New Synthesis in Moral Psychology. *Science*, **316**, 998-1002.
- Johansson, P., Hall, L., Sikström, S., and Olsson, A. (2005) Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, 310, 7 OCTOBER, 116-119.
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Marc Hauser, M., and Damasio, A. (2007) Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, 446,19April, 908-911.
- Koenigs, M. and Tranel, D. (2007) Irrational Economic Decision-Making after Ventromedial Prefrontal Damage: Evidence from the Ultimatum Game. *The Journal of Neuroscience*, 27(4), 951–956.
- Moll, J., Eslinger, P. J., de Oliveira-Souza, R. (2001) Frontopolar and anterior temporal cortex activation in a moral judgment task: Preliminary functional MRI results in normal subjects. *Arq Neuropsiquiatr*, 59, 657-664
- Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F. and Grafman, J. (2005a) Opinion: The neural basis of human moral cognition. *Nature Reviews Neuroscience*, 6(10), 799-809.
- Moll, J., de Oliveila-Souza, R., Moll, R. T., Ignácio, F. A., Bramati, I. E., Caparelli-Dáquer, E. M., Elsinger, P. J. (2005b) The moral affiliations of disgust: a functional MRI study. *Cognitive & Behavioral Neurology*,18(1), 68-78.
- Moll, J., de Oliveila-Souza, R., Zahn, R. (2008) The Neural Basis of Moral Cognition: Sentiments, Concepts, and Values. *Annals of the New York Academy of Sciences*, 1124, 161-180.


Thank you for your attention