

HUME'S THEORY OF MORAL IMAGINATION

Mark Collier

David Hume endorses three statements that are difficult to reconcile: (1) sympathy with those in distress is sufficient to produce compassion toward their plight, (2) adopting the moral point of view often requires us to sympathize with the pain and suffering of distant strangers, but (3) our care and concern is limited to those in our close circle. Hume manages to resolve this tension, however, by distinguishing two types of sympathy. We feel compassion toward those we perceive to be in distress because *associative* sympathy leads us to mirror their emotions, but our ability to enter into the afflictions of distant strangers involves *cognitive* sympathy and merely requires us to reflect on how we would feel in their shoes. This hybrid theory of sympathy receives a good deal of support from recent work on affective mirroring and cognitive pretense. Hume's account should appeal to contemporary researchers, therefore, who are interested in the nature of moral imagination.

1. SYMPATHY AND COMPASSION

In *A Treatise of Human Nature* 2.2.7, "Of Compassion," Hume attempts to understand why we care about the pain and suffering of others. When we encounter agents in distress, he observes, we feel concern for their plight. But why do their sorrows and afflictions matter to us? Hume maintains that our compassionate responses to those in distress are "easy to explain" in terms of the principles of sympathy (T 2.2.7.2, SBN 369).¹ Human beings have a natural tendency, on his account, to mirror the emotional states of others. We care about their welfare, in other words, because we vicariously feel their pain.

Hume's surgery example provides a helpful illustration of this propensity to resonate with the affections of those around us:

Were I present at any of the more terrible operations of surgery, 'tis certain, that even before it begun, the preparation of the instruments, the laying of the bandages in order, the heating of the irons, with all

the signs of anxiety and concern in the patient and assistants, would have a great effect upon my mind, and excite the strongest sentiments of pity and terror. (T 3.3.1.7, SBN 576)

Spectators present at an operation, especially in 1739, would become alarmed. This secondhand response, however, is rather odd. It is obvious why the *patient* would become terrified as the doctor heats the irons and sharpens his instruments. But why would *spectators* feel this way? They are not the ones, after all, about to undergo surgery.

This difficulty can be resolved, according to Hume, in terms of emotional communication. Spectators in the operating room become terrified, on his account, because the fear and anxiety of the patient are transmitted to them: “As in strings equally wound up, the motion of one communicates itself to the rest; so all the affections readily pass from one person to another, and beget correspondent movements in every human creature” (*ibid.*). Hume refers to this transfer of affect, at one point, as a species of “contagion” (T 3.3.3.5, SBN 605). But talk of emotional contagion, of course, is entirely metaphorical. Viruses can spread through physical contact. But how does one literally catch someone else’s emotions?

Hume maintains that the mystery surrounding emotional contagion can be dispelled at the level of associations. The process begins when spectators make attributions concerning their feelings to others. We cannot directly observe the contents of other minds, of course, but we can nevertheless infer them on the basis of sensory cues (T 3.3.1.7, SBN 576). We recognize that those who wince and moan are in a state of pain, for example, because this feeling is associated with such gestures and exclamations. Our *idea* of their pain, moreover, is subsequently transformed into an *impression*²:

When any affection is infus’d by sympathy, it is at first known only by its effects, and by those external signs in the countenance and conversation, which convey an idea of it. This idea is presently converted into an impression, and acquires such a degree of force and vivacity, as to become the very passion itself, and produce an equal emotion, as any original affection. (T 2.1.11.3, SBN 317; cf. T 3.3.1.7, SBN 576)

Hume’s official position is that ideas and impressions differ solely in terms of their force and vivacity (T 2.1.11.7, SBN 318–19; cf. T 1.1.1.3, SBN 2). When ideas are sufficiently enlivened, therefore, this distinction would collapse.³

The feelings of others, it seems, rub off on us. We admire the lifestyles of the rich and famous, for example, because we “enter into” and “partake of their pleasure” (T 2.2.5.14, SBN 362). But our tendency toward affective mirroring is on clearest display, according to Hume, when it

comes to negatively valenced emotions. The afflictions and sorrows of others, as he puts it, "have always a stronger and more lasting influence than any pleasure or enjoyment" (T 2.2.7.2, SBN 369). We feel compassion toward those we perceive to be in distress, therefore, because our ideas of their suffering are especially lively and thus easily converted into impressions.⁴

Hume remarks that "the minds of men are mirrors to one another" insofar as they "reflect each other's emotions" (T 2.2.6.21, SBN 365). This observation was, of course, remarkably prescient. Social neuroscientists have recently discovered the existence of *affective mirror systems* in the brain that enable us to feel the pain of others.⁵ These researchers have developed a novel paradigm with which to study our responses to the emotional states of others. In these experiments, brain-imaging techniques are used to compare neural activity in two conditions: (1) when one feels pain and (2) when one observes another in pain. Singer et al. (2004) measured activation in the "pain matrix" of the brain, for example, as stimulation was applied to the hands of subjects via electrodes; they subsequently found similar firing patterns when subjects were presented with signs that those dear to them were in pain.

The Singer experiments involved emotional communication between loved ones. But further studies have shown that this propensity to reflect the emotions of others extends to strangers as well.⁶ These studies lend support, moreover, for Hume's associative hypothesis. Why is it that the same neural circuits fire when we *feel* pain as when we *observe* pain in others? An influential recent proposal is that mirror systems get wired up through Hebbian learning.⁷ Neurons that fire together, on this account, wire together. The fact that parents imitate the facial expressions of their children in pain is sufficient to explain, therefore, how neural connections are established between the neural networks that process firsthand and secondhand pain.⁸

2. SYMPATHY AND MORAL JUDGMENT

In *Treatise* 3.3.1., "Of the origin of the natural virtues and vices," Hume attempts to understand why we make moral judgments in situations where our interests are not at stake. It might be obvious why we praise beneficent characters, for example, when they advance our personal ends. But why do we feel the same way toward those who promote the good of others? In order to make sense of these evaluations, Hume invokes our natural tendency to mirror the emotions of those around us: "[It is] that principle, which takes us so far out of ourselves, as to give us the pleasure or uneasiness in the characters of others, as if they had a tendency to our own advantage or loss" (T 3.3.1.11, SBN 579; cf.

T 3.3.1.25, SBN 588–89). We feel esteem toward beneficent agents, in other words, because we resonate with the happiness of those who are helped. When we judge that such characters are virtuous, on Hume’s account, we express these vicarious feelings.

Hume recognizes that emotional contagion, however, cannot provide a complete account of moral judgment. The problem is that this propensity is based on principles of associations that give rise to significant biases in our moral thought.⁹ The *principle of resemblance* leads us to enter more easily into the feelings of those who share our language, manners, or professions (T 2.1.11.5, SBN 318). The *principle of causation* renders the affections of our friends and family more lively than those of strangers (T 2.1.11.6, SBN 318). But it is the *principle of contiguity*, perhaps, that has the most dramatic impact on our welfare considerations: “The breaking of a mirror gives us more concern when at home, than the burning of a house, when abroad, and some hundred leagues distant” (T 2.3.7.3, SBN 429). Emotional contagion, it seems, requires exposure. We simply do not resonate with the pain and suffering of others, as Hume puts it, without “sight of the object” (T 2.2.7.4, SBN 370; cf. 2.2.1.11.6, SBN 318).¹⁰

Our considered moral judgments, however, do not display these vicissitudes. We do not believe upon reflection that beneficent characters are any less virtuous, for example, because they happen to have been born in remote times or places. Virtues and vices, we ordinarily think, are distance-invariant:

We sympathize more with persons contiguous to us than with persons remote to us: With our acquaintance, than with strangers; With our countrymen, than with foreigners. But notwithstanding this variation of our sympathy we give the same approbation to the same moral qualities in *China* as in *England*. They appear equally virtuous, and recommend themselves equally to the esteem of a judicious spectator. The sympathy varies without a variation in our esteem. Our esteem, therefore, proceeds not from sympathy. (T 3.3.1.14, SBN 581)

Our moral evaluations do not fluctuate in accordance with our personal relationships with agents. It is difficult to understand how these judgments, therefore, could depend on the principles of association. It is these principles that lead us, after all, to discount the welfare of those unrelated to us.

Moral judgments require us, as Hume puts it, to adopt a “general point of view.”¹¹ When we evaluate the merits of distant characters, we must set our personal interests and attachments to the side and “place ourselves” in the position of those close to them (T 3.3.1.15, SBN 582; cf. T 3.3.3.2, SBN 602–3).

Tis therefore from the influence of characters and qualities, upon those who have an intercourse with any person that we blame or praise him. We consider not whether the persons, affected by the qualities, be our acquaintance or strangers, countrymen or foreigners. (T 3.3.1.17, SBN 582)

The moral point of view takes into consideration the feelings of those who interact with the agent. It incorporates, as Philip Mercer puts it, the “hypothetical pain or pleasure of hypothetical associates of the agent.”¹²

Consider our moral appraisal of Nero, for example, who murdered his own mother. We would not feel an intense hatred toward his cruelty, on Hume’s account, because this event took place in such a remote time and place. Indeed, we might feel a more lively resentment toward a bully who is cruel to children in our neighborhood. But we do not really think that one who steals lunch money is more villainous than one who sings while his nation burns. We recognize that we would feel greater disgust toward Nero if we lived under his bloody rule (T 3.3.1.16, SBN 582).

We blame equally a bad action, which we read of in history, with one perform’d in our neighborhood t’other day: The meaning of which is, that we know from reflexion, that the former action wou’d excite as strong sentiments of disapprobation as the latter, were it plac’d in the same position. (T 3.3.1.18, SBN 584)

Our capacity to reflect on how we would feel in hypothetical scenarios, according to Hume, enables us to “correct” the biases in our prereflective moral thought (T 3.3.1.17, SBN 583). Our considered moral judgments, on his account, express these adjusted sentiments.

3. A PUZZLE IN HUME’S ETHICAL THEORY

Hume recognizes that this solution, however, appears to “contradict” a crucial assumption of his theory of justice (T 3.3.1.16, SBN 586). His argument for the artificiality of justice, after all, presupposes that human beings are indifferent toward the welfare of those outside our close circle (T 3.2.2.8, SBN 488). Limited benevolence, as John Rawls would put it, is a circumstance of justice. Our ancestors were required to invent justice conventions because they understood that strangers bear “no kindness” toward them and thus could not be trusted to refrain from their possessions or reciprocate in social exchanges (T 3.2.5.8, SBN 520–21). These conventions would be superfluous, therefore, if we had the capacity to sympathize with the feelings of distant strangers. Agents with a “tender regard” for one another, as Hume puts it, would have little use for “distinctions and limits of property and possession” (T 3.2.2.16, SBN 494; cf. EPM 3.6, SBN 184–85).

Hume refers to our capacity to enter into the feelings of distant characters as “extensive sympathy” (T 3.3.1.16, SBN 586, cf. T 3.3.6.3, SBN 619). But the problem is that extensive sympathy with those in remote times or places, on his account, would produce extensive concern. We can express this tension in terms of a logically inconsistent triad. It appears that Hume endorses each of the following statements:

- (1) Sympathy with the pain and suffering of others is sufficient to produce lively feelings of concern toward their plight. (S→C)
- (2) We cannot make moral judgments about the cruelty of distant characters unless we sympathize with the pain and suffering of those in their close circle. (M→S)
- (3) We can make moral judgments about the cruelty of distant characters, even though we feel little or no concern toward the welfare of those in their close circle. (M&~C)

If any two of these propositions is true, however, the third must be false. Hume simply cannot, as he seems to do, endorse all three.

Hume maintains that the appearance of inconsistency can be “easily removed” by carefully examining the psychological foundations of our moral judgments (T 3.3.1.16, SBN 586). The crucial point is that our judgments concerning distant characters are based on counterfactual reasoning. When we judge that these agents are virtuous or vicious, according to Hume, we are asserting that we *would* feel love or hatred if we were close to them. We are not saying, however, that we *actually* feel this way: “Sentiments must touch the heart, to make them controul our passions: But they need not extend beyond the imagination to make them influence our taste” (ibid.). Adopting the general point of view, then, has little impact on our sentiments. Our sympathy with those in remote times or places, as he puts it, is “far from being as lively as when our own interest is concern’d, or that of our particular friends; nor has it such an influence on our love and hatred” (T 3.3.1.18, SBN 584; cf. T 3.3.1.21, SBN 585). Hume appears to have found a way, then, to drive a wedge between extensive sympathy and limited benevolence. Moral reflection changes the way we talk about distant characters, on his account, but it does not significantly alter the way we feel about them (T 3.3.1.21, SBN 585; cf. T 3.3.3.2, SBN 603).

This solution to the puzzle, however, is not yet complete. Let us grant that moral judgments express the sentiments that we would have in hypothetical scenarios. But this leaves open the question of how one can determine the truth-value of these counterfactuals. Hume proposes that we must rely on our imaginations as a guide. We must lay out the

scenario in our mind's eye, as it were, and envision how we would feel in these circumstances. His comparison with our aesthetic judgments about poorly written texts is highly instructive:

When I run over a book with my eye, I imagine that I hear it all; and also, by the force of imagination, enter into the uneasiness which the delivery of it would give the speaker. The uneasiness is not real; but as such a composition of words has a natural tendency to produce it, this is sufficient to affect the mind with a painful sentiment, and render the style harsh and disagreeable. (T 3.3.1.22, SBN 585–86)

Hume maintains that a “similar” operation of the mind underlies our moral judgments (T 3.3.1.22, SBN 586). We praise characters with benevolent dispositions, even if they are prevented from acting, because we imaginatively enter into the pleasures that they would produce if released. Along the same lines, we condemn the cruelty of distant characters because we enter into the pain and suffering of those in their close circle. But notice that this appears to take us back to where we started. If judging that Nero is vicious requires us to entertain ideas of the horrible torments of his victims, how could this possibly fail to produce lively feelings of compassion toward their plight?

The crucial point is that our moral judgments about distant characters are based on epistemic attitudes that are “inferior to *belief*” (T 3.3.1.20, SBN 585). When we imagine the afflictions of those close to Nero, for example, we never lose sight of the fact that we are actually safe in our study. Our sentimental responses are much weaker than they would be, therefore, if we genuinely believed that our real interests were at stake (T 3.3.1.18, SBN 583–84; cf. T 3.3.1.30, SBN 591). Compare our aesthetic judgments concerning a building that *appears* to be dangerous, even though we are fully assured of its safety. We say that this structure is disagreeable on the grounds that it produces in us a “kind of fear”; but this response pales in comparison to our intense fright when we stand under a wall that we “really think tottering and insecure” (T 3.3.1.23, SBN 586).

Contemporary research on the imagination allows us to develop this proposal.¹³ Adopting the moral point of view, in current terms, involves “cognitive pretence.”¹⁴ To understand how we would feel if we were closely related to Nero, we must represent a series of counterfactual propositions (for example, “I am a Roman citizen” “the year is 64 CE,” and such) in our “possible world box.” Even though we do not sincerely believe these statements, we can make inferences about the sentiments that would be appropriate if they were true. This computational model provides support, moreover, for Hume’s solution to the puzzle. The sentiments that

we experience during pretense episodes are “cognitively quarantined” from our practical reasoning system. Subjects who engage in pretence do not completely lose sight of reality. Children who pretend that a banana is a telephone, for example, do not really expect others to pick up their calls. Similarly, those who imagine that they are Roman citizens living under Nero’s tyrannical rule do not actually fear for their lives.

4. FIRST OBJECTION: IMAGINATION AND AFFECT

Our moral judgments about distant characters produce weak emotional responses, on Hume’s account, because we do not genuinely believe that we are closely related to them. One might wonder whether this solution to the puzzle is compatible, however, with his repeated assertions that the imagination has the power—independently of belief—to generate strong affective responses. Hume acknowledges that hypochondriacs make themselves sick, for example, by simply thinking that this is the case (T 2.1.11.7, SBN 319). But if the imagination has the power to make us ill, one might object, why cannot it generate vivid pangs of compassion? Hume could respond that there is an important difference between our epistemic attitudes in these cases. Hypochondriacs sincerely believe that they are sick; when we make moral judgments about Nero, however, we do not really believe (even for a second) that we are citizens of the Roman Empire.

It would appear that Hume’s solution conflicts, however, with his official views concerning our emotional responses to the arts. Consider the case of tragedy. Hume acknowledges that spectators feel compassion toward those who suffer on stage, even though their misfortunes are known to be “fictitious” (T 2.2.7.3, SBN 369; cf. T 1.3.10.7, SBN 122):

The whole art of the poet is employed, in rousing and supporting the compassion and indignation, the anxiety and resentment of his audience. They are pleased in proportion as they are afflicted, and never are so happy as when they employ tears, sobs, and cries to give vent to their sorrow, and relieve their heart, swoln with the tenderest sympathy and compassion. (OT 217)

Indeed, it is precisely the fact that audiences enter so deeply into the sorrows of these characters that makes it puzzling that we find tragedies so appealing.

Hume can explain our strong affective responses to theatrical tragedies, however, in terms of our proximity to the events that unfold. We feel compassion toward those on stage because the emotions that they display are highly contagious. The actor who plays Lear does not really grieve at the death of Cordelia, but he acts as if he does, and audiences

resonate with his gestures and expressions (T 2.2.7.3, SBN 369; cf. EPM 5.26, SBN 221–22). Our emotional responses to theatrical tragedies are softened, moreover, by the recognition that they are “nothing but a fiction” (OT 219; cf. T 1.3.9.15, SBN 115). Tragic poets borrow from historical figures and scenes to give their work verisimilitude (T 1.3.10.6; SBN 121–22). But audiences do not confuse fiction and reality; our affective responses to works of art might resemble those of real life, as Hume points out, but we can “easily distinguish” one from the other (T 1.3.10.10, SBN 630).

It would appear more difficult, perhaps, to reconcile Hume's solution with his iron-cage example. Those who are suspended from a high tower would succumb to fear, he proposes, even if they were fully assured that the bars of the cage would prevent their fall:

A man, who being hung out from a high tower in a cage of iron cannot forebear trembling, when he surveys the precipice below him, tho' he knows himself to be perfectly secure from falling, by his experience of the solidity of the iron, which supports him. (T 1.3.13.10, SBN 148)

We often *really* tremble, as Hume puts it, in response to “*imaginary* danger” (EPM 5.14, SBN 217). It would appear that sincere belief is not, then, a necessary condition of strong emotional responses.

This is an exceptional case, however, that can be explained away. The man in the cage is frightened, even though he believes that he is safe. But this is because his vivacious sensory impressions are associated with, and thus serve to enliven, his related ideas of descent and death (T 1.3.13.10, SBN 148). The imagination, in other words, is “here assisted by the presence of a striking object” (EPM 5.14, SBN 217; cf. T 2.3.9.23, SBN 445). But this is not the case with the general point of view. When we reflect upon our possible feelings in hypothetical scenarios, there are no perceptual cues that could intensify our passions. We do not witness the wails and moans of Nero's victims; as a result, our ideas of their afflictions are not converted into lively impressions.

5. SECOND OBJECTION: SYMPATHY AND SIMULATION

To condemn the cruelty of Nero, on Hume's account, we must take into consideration the pain and suffering of his victims. This moral judgment produces little or no concern toward their plight, however, because we do not sincerely believe the assumptions on which it is based. Sympathy only produces compassion when we *perceive* the pain and suffering of others, but adopting the moral point of view merely requires us to *imagine* ourselves as spectators of their afflictions. This proposal can be developed, as we have seen, in terms of cognitive theories of pretense. But

there are alternative models of pretense available in cognitive sciences; these accounts, moreover, are quasi-perceptual in nature. One might reasonably wonder, then, whether they are compatible with Hume's solution to the puzzle.

Consider "simulation" theories of imagination.¹⁵ We identify the emotional states of others, on these accounts, by *reenacting* or *recreating* the contents of their beliefs and desires. These "proxy" states are fed as input into our affective response system, and the output is used to make inferences about how they feel (Goldman 2006, 170).¹⁶ Let us clarify this proposal with a simple example: D. Kahneman and A. Tversky's airport experiment.¹⁷ Participants were asked to identify the feelings of two characters, Mr. Tees and Mr. Crane, whose limousine arrives at the airport thirty minutes after their flights were scheduled to depart. Mr. Crane is informed that his plane left on time; Mr. Tees is told that his flight was delayed and recently took off. Ninety-six percent of the participants who reflected on this imaginary scenario reported that Mr. Tees would have been more annoyed. How did they make this attribution? It is unlikely that they consulted folk theories about the feelings that are appropriate in this situation. It is more plausible to think that they envisioned themselves in similar circumstances and used the results of this imaginative procedure as a predictive tool.

Simulation theory provides us with a novel formulation of the general point of view. To make judgments about distant characters, on this account, one must attempt to recreate the feelings and attitudes of those in their close circle. Moral distinctions would depend on whether this procedure generates sentiments of love or hatred, resentment or gratitude. Simulation theory sheds further light on why it is, moreover, that moral judgments produce weak affective responses. When we simulate the feelings of others—unless we are seriously delusional—we must "bracket" our own mental states. Our psychological states, as it is said, are taken "off-line." Consider the airport experiment once again. We can grant that participants relied on mental simulations to make predictions about the feelings of Mr. Tees. But it does not follow that those who made such attributions were ever seriously annoyed. And the reason for this is clear: simulators remain cognizant of the fact that *they* are not the ones who missed their flight and that the feelings of vexation properly belong to the target.

One might object that this defense of Hume's solution to the puzzle is based on an impoverished notion of perspective taking. Adopting the general point of view involves imagining how we would feel if we found ourselves in the situation of others. But what if this imaginary procedure was more radical in nature? In particular, what if it requires

us, as Robert Gordon has suggested, to *trade places* with others and imagine how *they* feel in these circumstances?¹⁸ Projecting ourselves into the circumstances of distant characters does not produce strong affective responses, according to Hume, because we do not lose sight of our actual interests. But what if spectators underwent an “egocentric shift” and viewed the world from the eyes of others?

Consider Gordon's interpretation of the airport experiment. To identify the feelings of Tees or Crane, he proposes, simulators must make a series of adjustments, such that the indexical “I” is now occupied by the target (Gordon 1995a, 734).

[O]nce a personal transformation has been accomplished, there is no remaining task of mentally transferring a state from one person to another, no question of comparing Mr Tees to myself. For insofar as I have recentered my egocentric map on Mr Tees, I am not considering what RMG would do, think, want, and feel in the situation. Within the context of the simulation, RMG is out of the picture altogether. (Gordon 1995b, 56)

The crucial point is that we leave ourselves behind when we step into the shoes of others. We do not analogically infer from what *we* would feel to what *they* would feel. Just as good method actors transform themselves into their characters, simulators do not merely pretend that they find themselves in the situation of others; rather, they *identify* themselves as the subjects of these states. It is difficult to see how this variety of simulation could fail to have a profound emotional response: if the line between pretense and reality were blurred, our simulated feelings would go “on-line” (Gordon 1995a, 739).

Hume can offer a decisive reply, however, to this line of objection. Consider the related challenge that he offers to egoists who attempt to reduce all moral judgments to considerations of selfish advantage. Hume rejects such proposals on the grounds that they cannot account for our praise and blame of beneficent characters in remote times or places (EPM 5.7–10, SBN 215–16). It is implausible that *spatially* remote agents will promote our interests; and it is impossible that *temporally* distant ones can do so. Hume recognizes that a sophisticated egoist might respond, however, that we make these judgments by imaginatively projecting ourselves into the shoes of those who receive benefits. When we evaluate distant characters, on this proposal, we express the happiness that we would feel in these circumstances.

Hume presents the sophisticated egoist, however, with a serious dilemma. When we imagine ourselves in the shoes of others, we either distinguish reality and pretense, or else we confuse them. But either

way, the egoist cannot explain our moral judgments about distant characters. If we do not sincerely believe that our interests are affected by these agents, then we would not feel esteem toward them. It is hard to understand, as Hume puts it, how a “*real* sentiment or passion can ever arise from a known *imaginary* interest, especially when our real interest is still kept in view” (EPM 5.13, SBN 217). But it is just as difficult to grasp why we would praise the generosity of distant characters if we confuse reality and pretense and somehow transform ourselves into members of their close circle:

No force of imagination can convert us into another person, and make us fancy, that we, being that person, reap benefit from those valuable qualities, which belong to him. Or if it did, no celerity of imagination could immediately transport us back, into ourselves, and make us love and esteem the person, as different from us. Views and sentiments, so opposite to known truth and to each other, could never have place, at the same time, in the same person. (EPM 6.3, SBN 234)

The challenge for egoists is not to explain, in other words, why *someone else* approves of this agent; rather, it is to account for the fact that they do so. But this means that they cannot completely leave their own perspective behind when they enter into the feelings of others.

The egoist simply cannot have it both ways. And neither can the radical simulation theorist. If we trade places with others when we adopt the general point of view, as Gordon would have it, then *we* would no longer be the subject of these judgments. When we imagine ourselves in the shoes of others, in other words, we must remain cognizant of the fact that we are engaged in pretense. Hume’s solution to the puzzle is compatible, therefore, with both moderate and radical versions of simulation theory. If we distinguish between reality and pretense, as Goldman would have it, then the simulation procedure would produce relatively weak affective responses; if we confuse reality and pretense, as Gordon would have it, then the simulation might produce lively feelings, but they would no longer be ours.

One might object that Hume is hung by his own petard. When we make moral judgments, on his account, we do not sincerely believe the assumptions on which they are based. But then it becomes difficult to understand how these *imaginary* relationships could influence our *real* sentiments. This criticism, however, misses the point entirely. Hume’s official position, as we have seen, is that the moral imagination does not significantly change our interpersonal attitudes. It should be noted that his position is identical, in this regard, to the one put forth by Adam Smith. Smith’s account of sympathy requires us to project ourselves

imaginatively into the shoes of others: we “enter as it were into his body, and become in some measure the same person with him.”¹⁹ But he makes it clear that spectators do not become confused about their true identity:

The thought of their own safety, the thought that they themselves are not really the sufferers, continually intrudes itself upon them; and though it does not hinder them from conceiving a passion somewhat analogous to what is felt by the sufferer, hinders them from conceiving anything that approaches the same degree of violence. (Smith 2002, 27)

When we adopt the perspective of others, according to Smith, we do not blur the boundary between self and other. Adopting the moral point of view does not produce strong affective responses, then, because we never lose sight of the fact that we are engaged in role taking.

6. THIRD OBJECTION: MORALITY AND THE WILL

Adopting the moral point of view, according to Hume, requires us to imagine how we would feel in hypothetical scenarios. As Barry Stroud points out, however, this proposal appears to sever the conceptual link between morality and the will:

Hume is in danger of breaking the connection between moral judgments and the will by arguing simply that feelings are involved or alluded to in some way or other in the making of every moral judgment. He must show precisely how a “disinterested” moral judgment is nevertheless “active”, or how a thought or belief about merely possible feelings can lead us to act. That is something he never explains.²⁰

Hume criticizes the moral rationalists on the grounds that they cannot account for the push and pull of moral considerations (T 3.1.1.6, SBN 457). But it now seems that Hume's account does not fare any better. The practical force of morality would be easy to explain if our evaluations were constituted by *actual* feelings.²¹ But moral motivation is much harder to understand if our judgments are based on our *possible* feelings in counterfactual circumstances.

It should be noted that this difficulty arises for any “spectator” theory of morals. The fact that impartial *observers* approve of a particular course of action, as Gilbert Harman points out, does not entail that *agents* must feel this way.²² This problem can be stated most forcefully, perhaps, in terms of simulation theory. To adopt the moral point of view, on the Humean account, we must “bracket” our psychological states. But if moral judgments require us to take our own beliefs and desires

“off-line” and substitute those of others, it becomes extremely difficult to see how they could influence the will. How could these judgments be *action-guiding* if they represent *someone else’s* mental states?

Let us consider a simple example. Suppose that we stand to gain a significant fortune from acts of wanton cruelty. Would the judgment that these actions are morally wrong influence our practical deliberation? It is hard to see, on Hume’s account, how it could. To adopt the moral point of view with regard to our own conduct, we must imagine the feelings of those who lack our particular interests. But this judgment is based on epistemic attitudes that fall short of belief; as a result, it seems that they will not significantly influence our passions. When we judge that our actions are morally wrong, on Hume’s account, we are asserting that we *would* disapprove of them *if* we were disinterested. But this says nothing about our actual mindset. It might be the case that, as a matter of fact, we remain entirely partial to our interests.

To account for the practical force of morality, then, Hume must explain why the feelings of disinterested spectators matter to us. But he has the theoretical resources to do so. When we adopt the general point of view, on his account, we discover that others would not go along with our selfish behavior. This information does not, on its own, influence our passions. These considerations will only move us if we are responsive to the attitudes of others. Hume’s theory of sympathy, however, makes this inevitable. We do not only affectively resonate, on his account, with the pleasures and pains of others, but we also “naturally sympathize with others in the opinions that they entertain of us” (T 3.2.2.24, SBN 499; cf. T 2.1.11.9, SBN 320):

The sentiments of others can never affect us, but by becoming in some measure, our own, in which case they operate upon us, by opposing and increasing our passions, in the very same manner, as if they had been originally deriv’d from our own temper and disposition. (T 3.3.2.3, SBN 593)

The gap between the spectator and agent perspectives is closed, then, when we come to see ourselves as others would (Brown 1994, 31). We would no longer merely *imagine* that others feel resentment toward our vices; rather, we would *actually* hate ourselves when we contemplate these actions.

7. CONCLUSION

Sympathy plays two important roles in the *Treatise*. The first involves *compassion*: we care about those in distress because we vicariously feel their pain. The second involves *moral judgment*: we evaluate

distant characters by imagining the feelings of those in their narrow circle. Hume recognizes that assigning this double duty to sympathy, however, appears to undermine his theory of justice. His argument for the artificiality of justice presupposes that we have little compassion toward the welfare of strangers. But this is difficult to square with the claim that we can extensively sympathize with the feelings of those in distant times and places. It seems that we could not make moral judgments that are universal in scope, on his account, without expanding our circle of concern.

We can now see that this objection rests on an equivocation. The problem dissolves once we have disambiguated the two senses of sympathy at work in Hume's moral psychology:

- (1) *Associative* sympathy with the pain and suffering of others is sufficient to produce lively feelings of concern toward their plight.
- (2) We cannot make moral judgments about the cruelty of distant characters unless we *cognitively* sympathize with the pain and suffering of those in their close circle.
- (3) We can make moral judgments about the cruelty of distant characters, even though we feel little or no concern toward the welfare of those in their close circle.

This set of revised statements is entirely consistent. When we perceive someone in distress, our minds serve as mirrors, and we automatically feel that person's pain. But this is not true of the general point of view. When we imagine the feelings of those closely related to distant characters, we do not lose sight of the fact that our actual ties to these agents are weak; thus, our ideas of their sorrows and afflictions are not converted into lively impressions.

Hume's solution to the puzzle appears to have been on the right track. Recent work in social neuroscience has demonstrated that affective mirroring and cognitive perspective taking recruit distinct areas of the brain. Our capacity to resonate with the emotional states of those around us depends on mirror properties of neural systems in the emotional areas of the brain; but our capacity to imagine ourselves in the circumstances of others recruits cognitive areas in the prefrontal cortex.²³ Hume's hybrid theory of sympathy should appeal, then, to contemporary researchers who are interested in the nature of moral imagination.

University of Minnesota, Morris

NOTES

1. References to Hume's works are given in parentheses in the text: T = David Hume, *A Treatise of Hume Nature*, ed. David Fate Norton and Mary J. Norton (Oxford: Oxford University Press, 2002), cited by book, part, section, and paragraph number. SBN = David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2d ed., rev. P. H. Nidditch (Oxford: Clarendon Press, 1978), cited by page number.

EPM = David Hume, *Enquiries concerning Human Understanding*, ed. Tom L. Beauchamp (Oxford: Oxford University Press, 1999), cited by section and paragraph number. SBN = David Hume, *An Enquiry concerning Human Understanding and concerning the Principles of Morals*, ed. L. A. Selby-Bigge, 3rd ed., rev. P. H. Nidditch (Oxford: Clarendon Press, 1975), cited by page number.

OT = "Of Tragedy" in E. F. Miller, *Hume's Essays: Moral, Political, and Literary* (Indianapolis, IN: Liberty Fund, 1987).

2. Hume appears to regard this transformation process as uncontroversial. We commonly acknowledge that impressions are changed into ideas when they *lose* vivacity; so why should we deny that ideas are converted into impressions when they *gain* it? It should be pointed out, moreover, that Hume's distinction between ideas and impressions was never sharp. In T 1.1.1.1 (SBN 2), for example, he maintains that it breaks down in aberrant cases: madness or dreams enliven ideas to the point where they are transformed into impressions. Our capacity to sympathize with the distress of others demonstrates that this conversion process can take place in ordinary circumstances as well.

3. Ideas are transformed into impressions when they are sufficiently enlivened; but what supplies our ideas of the afflictions of others with the requisite force and vivacity? It is at this point that Hume's account takes its notorious turn: he locates the animating source in "impressions of ourselves" (T 2.1.11.4, SBN 317): "The stronger the relation is betwixt ourselves and any object, the more easily does the imagination make the transition, and convey to the related idea the vivacity of conception, with which we always form the idea of our own person" (T 2.1.11.5, SBN 318). These impressions contain maximal degrees of force and vivacity; as a result, they serve to intensify any idea connected to them. If we associate ourselves closely enough with other agents, therefore, their affections will become psychologically real for us.

4. It is essential to compassion, on Hume's analysis, that one desire to alleviate the suffering of others (T 2.2.9.3, SBN 382). One must not only resonate with their distress, in other words, but one must also be disposed to help them.

Def. Compassion: A feels compassion towards B if and only if (i) A has a lively impression of B's pain and (ii) A has a prima facie desire to improve B's condition.

Hume maintains that the associative principles of sympathy, however, can account for this motivational component. If we came across someone asleep in

a field about to be trampled by horses, he writes, these principles would lead us to “immediately run to his assistance” (T 2.2.9.13, SBN 385). We come to the aid of those in distress, then, because we vicariously share their pain and suffering.

5. B. Wicker, C. Keysers, J. Pially, J. P. Royet, V. Gallese, and G. Rizzolatti, “Both of Us Disgusted in My Insula: The Common Neural Basis of Seeing and Feeling Disgust,” *Neuron* 40 (2003): 655–64, at 661; T. Singer, B. Seymour, J. O’Doherty, H. Kaube, R. J. Dolan, and C. D. Frith, “Empathy for Pain Involves the Affective But Not Sensory Components of Pain,” *Science* 303 (2004): 1157–62, at 1159; and T. Singer, “The Neural Basis and Ontogeny of Empathy and Mind Reading: Review of Literature and Implications for Future Research,” *Neuroscience and Biobehavioral Reviews* 30 (2006): 855–63, at 858.

6. I. Morrison, D. Lloyd, G. di Pellegrino, and N. Roberts, “Vicarious Responses to Pain in Anterior Cingulate Cortex: Is Empathy a Multisensory Issue?” *Cognitive, Affective, and Behavioral Neuroscience* 4 (2004): 270–78, at 274; and M. Botvinick, A. P. Jha, L. M. Byllesma, S. A. Fabian, P. E. Solomon, and K. M. Prkachin, “Viewing Facial Expressions of Pain Engages Cortical Areas Involved in the Direct Experience of Pain,” *Neuroimage* 25 (2004): 312–19, at 317.

7. C. Keysers and D. Perrett, “Demystifying Social Cognition: A Hebbian Perspective,” *Trends in Cognitive Sciences* 8 (2004): 501–7, at 504; and M. Brass and C. Heyes, “Imitation: Is Cognitive Neuroscience Solving the Correspondence Problem?” *Trends in Cognitive Sciences* 9 (2005): 489–95, at 491.

8. C. Keysers and V. Gazzola, “Towards a Unifying Neural Theory of Social Cognition,” *Progress in Brain Research* 156 (2006): 379–401, at 392.

9. It should be noted that Hume stands on firm empirical ground here. There is a vast amount of evidence from social psychology that confirms the claim that we resonate more strongly with the affections of those related to us (see G. Lowenstein and D. Small, “The Scarecrow and the Tin Man: The Vicissitudes of Human Sympathy and Caring,” *Review of General Psychology* 11 [2007]: 112–26 for an overview). D. L. Krebs (“Empathy and Altruism,” *Journal of Personality and Social Psychology* 32 [1975], 1134–46) has demonstrated that the strength of our emotional responses to the electrical shock of others depends on our degrees of resemblance to them. B. Jones and H. Rachlin (“Social Discounting,” *Psychological Science* 17 [2006]: 283–86) have shown that our concern for the welfare of others is a direct function of our perceived social distance. There is also evidence to support Hume’s proposal, moreover, that we sympathize most strongly with those whose material interests are aligned with ours (T 2.2.9.6, SBN 383). J. T. Lanzetta and B. G. Englis (“Expectations of Cooperation and Competition and Their Effects on Observer’s Vicarious Emotional Responses,” *Journal of Personality and Social Psychology* 56 [1989]: 543–54) establish that the degree of our affective response to those in pain varies according to whether we stand in competitive or cooperative relations (551).

10. We associate ourselves with everyone else, according to Hume, because we resemble one another qua human being (T 2.2.7.2, SBN 369). But his official position is that this is not sufficient for compassion: it is only when the rela-

tion of resemblance is “united together” with contiguity or causation that we conceive the “sentiments or passions of others” in the strongest possible light and feel concern toward their plight (T 2.1.11.6, SBN 318).

11. There is not adequate space in this paper to address the important question of *why* it is that we adopt the general point of view. Hume’s official position is that we do so to avoid the contradictions that arise from our associative biases. There is an interesting disagreement among Hume scholars about how this claim should be interpreted. C. Korsgaard (“The General Point of View: Love and Moral Approval in Hume’s Ethics,” *Hume Studies* 25 [1999]: 3–41) maintains that the ultimate motivation for taking up the general point of view is to remove cognitive dissonance (at 25); G. Sayre-McCord (“On Why Hume’s General Point of View Isn’t Ideal—and Shouldn’t Be,” *Social Philosophy and Policy* 11 [1994]: 202–28) argues that we do so to avoid interpersonal conflict; R. Cohon (“The Common Point of View in Hume’s Ethics,” *Philosophy and Phenomenological Research* 67 [1997]: 827–50) maintains that we adopt this perspective to make more reliable predictions about which characters will produce love and hatred (at 843). This topic, of course, deserves a fuller treatment. In this paper, I am only concerned with the *how* question concerning the psychological mechanisms that enable us to adopt the moral point of view.

12. P. Mercer, *Sympathy and Ethics: A Study of the Relationship between Sympathy and Morality, with Special Reference to Hume’s Treatise* (Oxford: Clarendon Press, 1972), 56.

13. S. Nichols and S. P. Stich, *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds* (Oxford: Clarendon Press, 2003), 28–38.

14. The term “pretence” might suggest playfulness and frivolity. But this is far from the case. Our capacity to engage in pretense is anything but trivial in Hume’s moral psychology; indeed, this ability serves as the psychological foundations of moral development. Simply put, if we could not employ pretense, we would be trapped in our own “peculiar” or personal perspective, and we could not make disinterested moral judgments.

15. A. Goldman, *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading* (New York: Oxford University Press, 2006); and G. Currie and I. Ravenscroft, *Recreative Minds: Imagination in Philosophy and Psychology* (Oxford: Oxford University Press, 2002).

16. These pretend states *resemble* their actual counterparts, according to simulation theorists and thus produce similar emotional responses (Goldman 2006, 48–49, 284). Goldman recognizes that this proposal bears a close affinity to Hume’s famous dictum that ideas resemble impressions (150). When we close our eyes and imagine our room, Hume maintains, our ideas are (weaker) “exact representations” of our sensory impressions (T 1.1.1.3, SBN 3). But Goldman makes it clear that his analysis of similarity proceeds along different lines from Hume’s. Hume maintains that imaginary states resemble perceptual states in terms of their phenomenological force and vivacity, then, but Goldman explicates similarity in terms of neural or functional properties (151).

17. D. Kahneman and A. Tversky, "The Simulation Heuristic," in *Judgment under Uncertainty: Heuristics and Biases*, ed. P. Slovic, D. Kahneman, and A. Tversky (Cambridge: Cambridge University Press, 1982), 203. Cf. A. Goldman, "Empathy, Mind, and Morals," in *Proceedings and Addresses of the American Philosophical Association* 66 (1992): 17–41, at 21.

18. R. M. Gordon, "Sympathy, Simulation, and the Impartial Spectator," *Ethics* 105 (1995a): 727–42; and "Simulation without Introspection or Inference from Me to You," in *Mental Simulation: Evaluations and Applications*, ed. M. Davies and T. Stone (Oxford: Blackwell, 1995b), 53–67.

19. A. Smith, *The Theory of Moral Sentiments*, ed. K. Haakonsen (1759; Cambridge: Cambridge University Press, 2002).

20. B. Stroud, *Hume* (London: Routledge and Kegan Paul, 1977), 192. Cf. E. Radcliffe, "Hume on Motivating Sentiments, the General Point of View, and the Inculcation of 'Morality,'" *Hume Studies* 20 (1994): 39.

21. P. Foot, "Hume on Moral Judgment," *Virtue and Vices and Other Essays in Moral Philosophy* (Berkeley: University of California Press, 1978), 79.

22. G. Harman, *Explaining Value and Other Essays in Moral Philosophy* (Oxford: Clarendon Press, 2000), 184. Cf. C. Brown, "From Spectator to Agent: Hume's Theory of Obligation," *Hume Studies* 20 (1994): 19–35.

23. J. Decety and C. Lamm, "Human Empathy through the Lens of Social Neuroscience," *The Scientific World Journal* 6 (2006): 1155; and S. D. Preston and F.B.M. De Wall, "Empathy: Its Ultimate and Proximate Bases," *Behavioral and Brain Sciences* 25 (2002): 19.