

The illusion of biological variation: a minimalist approach to the mind

Open Talk and Discussion

Marc D. Hauser

M. Piattelli-Palmarini, J. Uriagereka and P. Salaburu (Eds.)

"Of Minds and Language: The Basque Country Encounter with Noam Chomsky"

Oxford University Press (in press)

Note: this is a transcript from a talk by M.D. Hauser

Initially, when I heard that I was going to be giving an open public lecture, I had a different audience in mind than the people who were here for the morning session. I have therefore been changing my talk as the day has gone on and hope that there is not too much repetition.

The topic that I want to talk about today falls under the title “The illusion of biological variation”. For those of you who have been staring at the image projected on the screen here, you may think that there is some kind of animation that is creating the motion. But that is a perceptual illusion: the image is completely static, with nothing moving at all, except that your visual system thinks it is. If you don’t believe me, focus on one of these concentric circles and look at the dot, and you will see that nothing is actually moving. Now, no matter how many times I tell you that the image is static, you won’t believe me — well, your visual system won’t believe me it can’t. Illusions are interesting because, no matter how aware we are of them, they simply won’t go away. Similarly, and by way of analogy, I will suggest today that much of the variation that we see in the natural world is in some sense an illusion because at a different level of granularity, there are some core, invariant mechanisms driving the variation.

As in any talk that attempts to go beyond one’s typical intellectual limits or comfort zone, I must first make a few apologies. The first one is to Chris Cherniak and other theoretical biologists, for my gross generalizations drawn from some of the very deep facts they have uncovered about the natural world. The second one is to Noam and other linguists because I am going to generate some wild speculations about language evolution from a very fragmentary bit of evidence. The third apology is to a class of philosophers, and in particular to John Rawls, for cutting out all the subtleties of argumentation that have gone on about utilitarianism on the one hand, and deontological

principles on the other, so that I can cut to the chase and tell you about how the moral faculty works. And then a final apology to John Cage and many minimalists in music and art, particularly for taking grotesque liberties with their theories and painting a slightly different picture of what I think they really were after.

The first point to make is that when we look upon the natural world, we immediately see extraordinary variation in animal forms, what looks like limitless variation, not just in size (from extremely small animals to immensely huge animals), but in shapes, material properties and so forth. Similarly, we see apparently limitless variation in the patterns of animal locomotion, including most noticeably, those observed in the air, on land and in the sea. Somebody raised a question earlier in the meeting about the immune system — again, a system with limitless variation in the kinds of responses that it generates to different kinds of problems in the environment. I want to call all of this observed variation, the “the illusion of biological variation.” It is an illusion, at least in part, because when biologists have looked deeply into the sources of variation in these different domains, as Cherniak’s talk illuminated this morning, we find something different — a common set of core mechanisms that generates the variation.

Let me put this into a historical context by quoting from two biologists who confronted the nature of biological variation. The first is Sewall Wright, who may be known to many of you. He was a distinguished evolutionary biologist who, following from Darwin, talked about the nature of adaptation and in particular the notion of an adaptive landscape. Here is what Wright pointed out in the 1930s, which I think is very telling in terms of the story I want to paint today. He says, “The older writers on evolution were often staggered by the seeming necessity of accounting for the evolution of fine details. Structure is never inherited as such, but merely types of structure under particular conditions.” Now, at the time Wright was discussing these matters, there were major revolutions afoot in genetics and molecular biology. If we fast-forward the story to today, here is a quote from Mark Kirschner, a systems biologist who makes very much the same point but takes it a little bit further and takes it in a direction that will hopefully have great appeal especially to the linguists in the audience who are interested in certain kinds of structural properties. Kirschner says, “Novelty in the organism’s physiology, anatomy or behavior arises mostly by the use of conserved processes in new combinations at different times, in different places and amounts, rather than by the invention of new processes.” This is very much in line with some of the things that Gabby [Dover] was saying this morning. “In the 4-5 billion years of cellular life on Earth, there have been four core processes leading to variation: rearrangement, repetition, magnification and division.” For those of you who have been tracking what has been happening in the minimalist program, you will see a kind of family resemblance to these four core processes.

Language

The idea that I want to push today — a project that Noam and I have been working on a bit over the last year or so, but I will take full responsibility for in terms of errors — is to invoke three principles that extend the minimalist program in linguistics to the mind and other domains of knowledge more generally. The first is that any time we observe an open-ended, limitlessly expressive, powerful system, it will be based on a fixed set of principles or mechanisms for generating the observed variation — i.e., some kind of generative, combinatorial system. Secondly, these generative mechanisms must in some sense interface with system-internal and -external processes, with nature potentially

finding the optimal solution given the current conditions. This allows for lots of accidental variation that happened before, but I will talk specifically about how it allows solutions to the current conditions. And then lastly, each variant we observe will be determined by some kind of process of pruning, where the local experience tunes up the biologically given options.

That is the idea in a nutshell. Now I want to take it a little bit further. The main idea in Noam's talk and alluded to in Gabby's talk, is that rather than having thousands and thousands of variants, we have one animal, one blueprint. I know that Gabby is somewhat opposed to this metaphor, but the idea is that we have some kind of conserved process that is generating all the variation. Of course one of the nicest stories to come out in the last twenty or so years is the account related to the Hox gene system, where we can see [in the projection (off-screen)] a direct mapping between genes that are basically building segmentation in the body forms of drosophila on top, and the mouse embryo below. The details that you see in between don't matter; the key point is the color coordination here that lines up patterns of segmentation that are being driven by evolutionarily ancient genetic mechanisms that have been conserved over evolutionary time. Part of the reason why this is important is because it has changed the nature of the way we think about the notion of homology. If you were simply to focus on anatomical form, and note how different things look, you would be missing the underlying genetic similarity that is extremely conserved and is homologous in that sense. The second is a recent paper by Bejan and Marden¹ which has received a lot of attention, especially by functional morphologists, claiming that all patterns of animal locomotion can be explained by equations relating force, energy, frequency and mass. And just to give you one example of a beautiful fit, consider the relationship [projection off-screen] between body mass on the X-axis and force on the Y-axis. Here we've plotted with one symbol all the animal species that move by running (mammals, reptiles and insects), all those that fly (birds, bats and insects), and those that swim (fish, marine mammals and crayfish). As you can see, they all align beautifully. Now much of the controversy surrounding these results is that experts in the field claim the analyses fail to account for complexities and variations observed. I am sure this is correct, but my guess is that the history of this debate will end up looking a lot like the debates in linguistics, where there are going to be some battles about the details, but what seems to be captured here are generalities.

The move that I want to make is that, given the kinds of depth of investigation that have gone on in biology over the last 30-40 or more years, when variation has in some sense been put to the side for the purpose of looking at explanatory mechanisms, there is a common theme that seems to keep emerging. What I want to ask now is basically whether that kind of move can be adopted in thinking about the nature of the human mind. Thus, for example, it certainly appears to be the case that there is limitless cultural variation. Can we account for it by some simple, primitive mechanisms, and then use pruning as a mechanism for selecting among the possible, biologically given variants? To test this question, we need to run the universal minimalist program of research. We first look for a core set of rules or mechanisms with a generative power of expression, interfacing with specific forms of knowledge. We next explore whether these mechanisms are present in other animals and the degree to which their presence in

¹ Bejan, A. & Marden, J.H. (2005). "Unifying constructal theory for scale effects in running, swimming and flying. *The Journal of Experimental Biology* 209, 238-248.

humans is unique to a domain of knowledge or more domain-general. Then we run the comparative analysis from genes to behavior, attempting to understand what limits the phenotypic space. Now here is a paradox to keep in mind. I don't think either Randy [Gallistel] or I want to be taken as saying that there is nothing unique about humans at all; the comparative evidence we have presented shows there are extraordinary abilities in animals and it is important to keep this in mind. Here, however, is the paradox I want to point to today. Gabby mentioned the genomics of humans, and of course one of the most interesting things about the study of genomics today is the fact that if you look at the genetic relationships or similarities between chimpanzees and humans, they are far more similar to each other than are chimpanzees and humans as a cluster to gorillas. Now that is surprising again if you think about their anatomy. Chimps look much more like gorillas than they do like human beings, and yet at the level of genetic similarity chimpanzees cluster with humans and not gorillas. That said, if we leap now from the anatomical level and genetic levels to the psychological level, we are faced with a fundamental problem. If we take some of the towering intellectual achievements in our history (and even some of the less towering intellectual achievements), the gap between us and them is extraordinary; in fact I would say it is larger than the gap we see between gorillas and chimpanzees on the one side, and the humble beetle on the other. So we have to somehow come to grips with the fact that the genetic level of similarity is not accounting for the psychological variation and differences we see.

So here is the outline for what I want to say in the rest of today's talk. I want to run through three examples. I am going to first come back to language in the way that I spoke about earlier today, and I will flesh out a little bit more of the argument and present some new data that bear on the conceptual richness in nonhuman animals. Then I will turn to some parallel arguments about the nature of the moral faculty. Then I will turn to music as another domain in which we can basically begin to ask similar kinds of questions, and finally I will end with some summary points about nature's solution to the various kinds of problems about variation and unity.

So first, language. As I described this morning, I am going to think about language as *a mind-internal computational system designed for thought and often externalized in communication, and as such, language evolved for internal thought and planning and only later was co-opted for communication*. What I want to do now is use this hypothesis as a wedge to pinpoint a disagreement in the literature which I think unfortunately is sort of missing the point. But I am going to use it as a way of showing some data that I think actually bear on the argument, capturing the difference between the way that Noam has talked about the internal computational system of language, and the way that Steve Pinker and others have talked about language as an adaptation for communication — a distinction that at some level is virtually impossible to resolve, because language is used for both functions, and the question of evolutionary origins is notoriously difficult especially for such a complicated trait.

Let us return to the FLB-FLN distinction that I raised this morning and that Cedric Boeckx picked up on in his talk. Let us begin to think about the ways in which understanding of what goes into FLB vs. FLN can help think about the nature of the evolutionary process vis-à-vis the internal computational system for thought and planning and its externalization in spoken language, or sign language. The hypothesis that Chomsky, Fitch and I have been pinned with is what Pinker and Jackendoff have

called “the recursion-only hypothesis”. But that is not actually what we said². What we said was that FLN — as an hypothesis — consists of the computations that enter into narrow syntax (we specifically spoke about recursion) and the interfaces to semantics and phonology. We think this is a useful way to frame the problem because it forces one to look not only at the evolution of the computational system alone, but also, how it interfaces with and is constrained by the other mind-internal systems. This move opens the door to interesting comparative issues, which I turn to next.

If we look at the songbirds, species that learn their vocalizations based on some innate structure that guides the process of acquisition with critical periods and windows of opportunity very much like language acquisition, what we see are exceptional capacities for vocal imitation. This is especially true in the open-ended song learners like starlings and lyrebirds and mockingbirds — very complex streaming together of sound patterns that looks like a rich combinatorial system; but no meaning. The variation that you see in the songbird system does not generate new meaning, it is simply “I’m-Fred-the-sparrow-from-New-York”. And we’re finished, that’s it. Change the variation a bit and it’s “I’m-Joe-from-California.” But I’m still a sparrow, and the meaning or function of my song is simple: I have a territory and I am looking for a mate. Fini! This is equally true of humpback whale song — again, very complicated, but the variation yields no new meaning. Now, here is a recording of a starling doing its own version of a goat and a chicken, as well as starling song material [plays recording]. Functionally, they use these songs in mate attraction, but they also use it as a sort of “No-vacancy” sign. They flood the habitat and say, “You don’t want to come here — there are goats and chickens and all sorts of other things around here. Don’t bother!” Now monkeys and apes, in striking contrast, show no evidence for vocal imitation. There is no capacity (and it has been fifty years of intensive looking by primatologists), absolutely no evidence for vocal imitation. Primates typically do not string their calls or notes together; no combinatorics evident at all, and weak meaning if at all in their vocalizations.

Thus far, most of what I have focused on concerns the sensory-motor side. Now I want to come back quickly to animal concepts. People talk about the Galilean revolution... I like “Gallisteleian” for talking about concepts... I think Randy has done a great deal to help the cause in thinking about animal concepts, especially in terms of the notion of isomorphism... For now, I want to focus on intentionality, and in particular, the puzzle concerning the richness of animal mental life and the poverty of their communicative expressions. It is what I have often described as the metamorphosis problem after Kafka’s story, in which Gregor Samsa, qua beetle, has profound thoughts about the world, but cannot convey them.

Let’s start with physiology, and mirror neurons in particular, and then build up to thought and behavior. In the mid-1990s, Giacomo Rizzolatti was recording from neurons in the pre-motor cortex of a macaque monkey when he noticed that cells firing in response to observing an experimenter grasp an object also fired when this same monkey grasped the same object, in the same way. Further recordings, from other cells in the premotor area, revealed a kind of gestural repertoire: cells firing when the action itself is in the repertoire of the animal, and the animal either performs the action or observes another individual performing the same action. There is a linking or coupling between perception and action. Now what I want to show you is how you can take these

² Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002). The faculty of language: What is it, who has it, and how does it evolve? *Science*, 298, 1569-1579.

physiological findings and look at how they may be instantiated in real world behavior in animals trying to make decisions about goals in the world and what they pay attention to when they perceive somebody acting in certain ways.

I am going to explain a study on the island of Cayo Santiago that was carried out by one of my terrific graduate students, Justin Wood. The star of the show, besides Justin, is the rhesus monkey again, and here is the experimental paradigm. You find an animal alone on the island. The monkeys on Cayo Santiago love coconut. Unfortunately, in over eighty years of living on this island, no single individual has ever figured out how to open one. Now this is a problem, because it is the most preferred food. If I crack open a coconut, they all come running. They call, they're very excited and so forth, but they can't work out how to open one on their own. However, they seem to understand that *we* can figure it out, so whenever we move towards some coconuts, they know that something interesting may happen, and it may be for them. This sets up a simple experiment. You show a subject two half coconuts face down. Because they are face down, but already cracked open, they can't see what is inside, but it is possible that one or both coconut halves has some flesh. For each experimental condition, the experimenter places these two half coconuts on the ground, face down, while the animal watches, and then approaches one of the coconut halves and interacts with it in some particular way before walking away. The psychologically relevant question is: does the particular form of interaction or action by the experimenter on the half coconut influence where the subject searches? The results I will now present [projected off-screen] reveal the proportion of subjects selecting the coconut acted upon by the experimenter. For each condition, we use one subject per trial, but multiple animals (between 20 - 24) per condition.

In the first condition, the experimenter simply grasped the top of the coconut but didn't lift it, and then walked away. Here, approximately 90% of the subjects approached the coconut that we grasped. Similarly, grasping the coconut with a pincer grip (i.e., index finger and thumb) also resulted in a selective approach to this coconut—approximately 85% of subjects. Interestingly, there are cells in the mirror neuron system that distinguish between a full hand grasp and a pincer grip; that is, cells that fire to a pincer grip do not fire to a full hand grasp, and vice versa. In a third condition, we grasped one coconut with a bare foot. Though the rhesus have never seen humans grasp in this particular way, rhesus will use their feet to pick up food, especially when they are hungry and attempt to carry as much food away as possible, using both hands, feet and their cheek pouches. In parallel with the first two conditions, 90% of the subjects go to the foot-grasped coconut. In the final condition in this series, we asked whether rhesus need to see the target goal in order to infer the subject's intentions or whether they can draw this inference when the goal is occluded or out of view; mirror neurons will fire when an agent reaches for and makes contact with a visible goal as well as when the goal is occluded. Similarly, rhesus selectively approach the occluded coconut when the experimenter reaches for it behind an occluder. Here, therefore, is a class of behaviors or actions that result in selective approaching behavior by rhesus. But there is a simple, and rather trivial explanation for all of these results: "Rhesus approach anything that an experimenter touches." If this is the rule they are following, then it is rather uninteresting, explained by simple associative mechanisms. If this is the proper interpretation, then any contact, intentional or not, with one coconut, should lead to selective approach. I turn next to conditions that directly explore the nature of the contact between experimenter and coconut.

In this next condition, the experimenter's hand merely flops on top of the coconut instead of grasping it. From a human perspective, it appears completely unintentional; the hand just flops on top of the coconut and then the experimenter walks away. In this condition, subjects show a 50-50 split: Some go to the box associated with the flop, the others go to the non-touched box. We also failed to get selective approach when the experimenter used a pair of pliers with a pincer grip, or contacted one coconut with a pole, or a machete; rhesus never use tools, have never seen the pliers or the pole, but have seen personnel on the island occasionally cracking open coconuts with a machete. The story that we are building up to, then, is not just about contacting or attending to the object, it is about the nature of the contact, intentional or not. Next, if you use the normal grasping mode with your hand, but you touch *next to*, as opposed to *on*, the coconut, they also show no preference. Interestingly, if you kneel down and grasp a coconut to use it to stand up (so you grasp it in exactly the same way but now it is just as a way of getting up), they also show no preference.

One of the arguments that has not yet been looked at in the mirror neuron story, but that we have begun to investigate, is whether rhesus distinguish between an action that is in the repertoire and therefore physically possible, as opposed to in the repertoire but irrational. For example, if I have a cup in front of me, I can reach for it by stretching out my hand, or I can reach for it by passing my arm under my leg — an odd gesture. Why would I do that? If an experimenter reaches between his legs and ends up in the hand-grasp position as the terminal position, our preliminary results fail to reveal a selective approach to the target box, even though the terminal state is both intentional and has the final grasping position. We are now in the midst of running a variant of this condition, one in which an experimenter holds a brick in each hand or has both hands empty, and then bends down and contacts the coconut with his mouth. If rhesus are like human infants similarly tested, they should contact the coconut with their mouth in the hands free condition but not in the hands with brick condition; that is, they should interpret the hands free condition as “if the experimenter has his hands free, but still contacts the coconut with this mouth, then there must be something important about using the mouth in this condition.” These studies, and others, suggest that there may be something like a large gestural repertoire that is being encoded for the agent's intentions, goals and the specific details of his or her action. Cross-cutting these dimensions may also be one that maps to rational vs. irrational trajectories vis-à-vis the end-goal state.

Now before I get too carried away with my excitement over these results, I want to make the following point in order to connect up with the final part of the language section. We seem to be uncovering, in both comparative studies and studies by developmental psychologists (such as my colleagues and collaborators Liz Spelke and Susan Carey, as well as others), what looks like an occasional mismatch between what individuals seem to know, on some version of knowing, and how they use that knowledge to act. So far what I have shown looks like a fairly good correspondence between their knowledge or attribution of knowledge, and their action, but now what I want to show you is an interesting mismatch. Back to Cayo Santiago and the rhesus monkeys. An experimenter finds a lone subject and shows this individual a table, indicating by tapping that it is solid; the experimenter then places one box on top of the table and a second box below the table, and then occludes the table and boxes; the experimenter then reveals an apple, holds it above the occluder, drops it, removes the occluder, and walks away, allowing the subject to search for the apple. Where do they go? To the bottom box, almost every single time. In fact, about 15% of the subjects look

in the bottom box and leave without ever checking the top box, as if they had decided that it must be in the bottom box, and thus, there is no point in checking the top box. Now we do the same experiment, but use the looking-time methodology that I talked about earlier. Here you remove the occluder and show that the apple is actually either in the top box or in the bottom box. Based on the search method that I just described, rhesus apparently expect to find the apple in the bottom box. Therefore, when it appears in the top box, rhesus should be surprised. From their perspective, this is a violation, so they should look longer when it appears in the top box than when it appears in the bottom box. But they don't. They look longer when the apple appears in the bottom box, which corresponds to a correct inference: that is, the apple can't appear in the bottom box as this would violate the physical principle of solidity. Thus, we see a dissociation between the knowledge that seems to be driving their looking responses as opposed to their searching behavior. We have a mismatch between the knowledge that seems to be driving a perceptual looking response, as opposed to a directed search response.

How can we tie this back into questions about the language faculty? Consider again the point I raised earlier concerning the possibility that the internal computations evolved for internal thought and then only subsequently, evolved further for the purpose of externalization in communication. What seems to be critically missing in nonhuman primates, and therefore primate evolution, is the interface between their rich conceptual system and the sensory-motor system, but most importantly, the system of vocal imitation. Monkeys and apes do not have the capacity for vocal imitation. As a result, they could never experience a lexical explosion. There is no way to pass the information on without vocal imitation. The implication here is significant. Independently of the story that emerges for the natural vocalizations of animals, and their putatively "referential" calls — such as the vervet monkeys' predator alarm calls — none of these systems show the kind of explosion in meaningful utterances that one sees in children from a very early age. This difference could have emerged for a variety of reasons, but one in particular is that there is no vocal imitation in nonhuman primates. If some genius vervet monkey invented an entire vocabulary of things for the environment, there would be no way to pass it on. It would just die with that individual. I think this argues very strongly for the idea that the system of thought was evolving for a very long time without any mechanism for externalization. For externalization to emerge, one species had to evolve the capacity to both link conceptual representations to distinctive sound structures, and for these structures to be passed on to others by means of imitation. Only one species seems to have worked this one out: *Homo sapiens*.

Morality

The same sort of questions arise for morality that arise for language, and interestingly we can think about the analogy between language and morality. I am certainly not the first to have made this kind of point, and let me just give a brief historical note. Noam mentioned several years ago, "Why does everyone take for granted that we don't learn to grow arms but rather are designed to grow arms? Similarly we should conclude that in the case of the development of moral systems, there is a biological endowment which in effect requires us to develop a system of moral judgment that has detailed applicability over an enormous range." The person who really picked this up in detail was the philosopher John Rawls, who in his 1971 classic, *A Theory of Justice*, made the following point: "A useful comparison here is with the problem of describing the

sense of grammaticalness.... There is no reason to assume that our sense of justice can be adequately characterized by familiar common-sense precepts...” [pp. 46-47] — very much like what we have been hearing over the course of the last couple of days about the linguistic moves and inventing of vocabulary — “or derived from the most obvious learning principles.” Again, one of the themes from today. “A correct account of moral capacities will certainly involve moral principles and theoretical constructions which go beyond the norms and standards cited in everyday life.”

Now this idea literally lay dormant for many many years. A few philosophers, Gil Harman, Susan Dwyer and most recently, John Mikhail, picked it up and began to argue for it more forcefully. Over the past three years, I have been exploring both the theoretical and empirical implications of the linguistic analogy with two fantastic graduate students of mine, Fiery Cushman and Liane Young; I realize that I probably shouldn’t wax so lyrical about my students, but, they really are as terrific as I claim! As a caveat before jumping into the empirical work, let me note that in striking contrast with the revolution in linguistics that took place 50 years ago, where there were already extremely detailed descriptions of language, there is nothing like this in the case of morality. Thus, we started our work with a significant deficit, especially with respect to achieving anything like descriptive adequacy. To start the ball rolling, we developed a website called the Moral Sense Test (moral.wjh.harvard.edu). It is a website that internet surfers visit on their own — if they have heard it discussed or if they Google “MST” (moral sense test), they will find us. Over a period of about two years we have collected data from approximately 100,000 subjects from 120 different countries, between the ages of 13 to 70. When an individual visits the site, he or she provides some biographical information — age, education, religious background, ethnicity, nationality and so forth — and then proceeds to read a series of moral dilemmas, followed by questions that ask about the permissibility, obligatoriness, or forbiddenness of an agent’s action. As an empirical starting point, we have made use of several artificial dilemmas created by moral philosophers to explore the nature of our intuitions concerning actions that involve some kind of harm. The use of artificial examples mirrors, in some ways, the artificial sentences created by linguists to get some purchase on the underlying principles that guide grammaticality, or in our case, ethicality judgments. Why go the route of artificiality when there are so many rich, real world examples in the moral domain?

The first reason that I need to spell out, though probably not as necessary with this audience as with many others, is that the use of artificial stimuli is a trademark of the cognitive sciences, providing a controlled environment to zoom in on the cognitive architecture of the human mind. A second reason, and I think more important in this particular context, is that real-world moral cases like abortion, euthanasia, organ donation, etc. have been so well rehearsed that our intuitions are gone. If I ask you “Is abortion right or wrong?”, you’ve got a view, and you’ve got a very principled view, in most cases. Whether I disagree with you or not is irrelevant. The main point is that you can articulate an explanation for why you think abortion is right or wrong. If you are interested in the nature of intuition, therefore, asking about real-world cases just won’t do. Our moral judgments are too rehearsed. Artificial cases are unfamiliar, but if we are careful, we can be manipulate them so that they capture some of the key ingredients of real world cases. What I mean by careful is that we set up a template for one kind of moral dilemma and then clone this dilemma, systematically manipulating only a key word or phrase in order to assess whether this small changes alters subjects’ moral

judgments. This method thus approximates a model in statistics or theoretical biology where one variable is manipulated while all others are held constant. Thus, for example, we take something like euthanasia, that relies in part on the distinction between actions and omissions, or more specifically, between killing and letting die, and then translate this into an artificial case such as the famous *trolley problems* that I will discuss in one moment. When philosophers make this move, they seem to be happy saying, “Well, my intuition tells me that this is right or wrong”. But for a biologically-minded, empirical scientist, this claim simply raises a second question: “Is the philosopher’s intuition shared by the man-on-the-street or is it a more educated decision?” This is an empirical question, and one that we can answer. Let me give you a flavor of how we, and others such as Mikhail and my new colleague at Harvard, Josh Greene, have begun to fill in the empirical gaps.

Consider four classic cases of the trolley problem. Somebody logs on to our website and they get some random collection of moral dilemmas. If they are trolley problems, they always begin with something like the following: A trolley is moving down a track when the conductor notices 5 people ahead on the track; he slams on the brakes but they fail, and he passes out unconscious; if the trolley continues on this track it will kill the 5 people ahead. Here is where the dilemmas begin to change. A bystander can flip a switch killing one person on a sidetrack, but saving the 5. And the question each subject will answer is, “Is it morally permissible to flip the switch?” When we ask people this question, 89% of our subjects say “Yes” to this question. Okay, now here is a small change in the problem. You are standing on a bridge, and you can push this fat guy off the bridge. He’s fat enough that he’ll stop the trolley in its tracks, but save the 5. You again ask “Is it morally permissible to push the fat guy?” Here, only 11% of subjects say “Yes”. Note that the utilitarians have a real problem here, because it is 1 vs. 5 in both cases, so if you are a utilitarian, you had better start looking for alternative explanations. Similarly, those with a deontological, non-consequentialist bent, are also in trouble because adhering to the rule that killing is wrong won’t work, as your actions result in the death of one in both cases.

Now the problem with these two dilemmas, looking at it scientifically, is that they have too many differences between them — there’s a fat guy, there’s a skinny guy, there are two tracks, there’s a redirection of threat, there is direct contact with a person vs indirect by means of a switch. What we need are cases where we reduce the variation leaving maybe only one principled distinction between the cases, enabling us look at the nature of the judgment. So here are two cases. The fat guy’s back, but now we have a loop on the track, and if you flip the switch, the train will go onto the loop, but then of course it comes back to hit the 5. However, the fat guy, who’s fat enough, can stop the trolley there and not kill the 5. You once again ask, “Is it morally permissible for the bystander to flip the switch?” The important thing to note here is that, just like the bridge case, this case can also be interpreted as using the man as an intended means for the greater good. If he’s not there, just flipping the switch does you no good, because the trolley goes on, comes back and kills the 5. The fat man (or in other versions, just a man with a heavy backpack) is the intended means, and your only hope for saving the five. Here, 52% of the people say that flipping the switch is morally permissible. Now note, in contrast to the bridge case which only generated an 11% permissibility judgment, there is a difference even though both use the intended means as a distinction. We’ll come back to this difference in a minute.

Here is a very similar case (bystander, loop, man) [projected off-screen], but now the man on the looped track is irrelevant because he's too thin to stop the trolley. However, in front of this man on the loop is a weight which is heavy enough to stop the trolley. This case can be interpreted as killing as a foreseen side-effect. Aiming for the man on the looped track makes no sense as he can't stop the trolley. Aiming for the weight makes sense as it can stop the trolley. Here, 76% of subjects say that it is morally permissible for the bystander to flip the switch, which is significantly greater than in the previous loop case. Importantly, these two cases involve impersonal harm, the trolley is redirected, there is only one man on the looped tracks, and the greater good is five saved in both. One of the potentially significant differences is between intended vs. foreseen harmful consequence. That is, using the man to save five as opposed to foreseeing the man's death to save five. .

These cases are just the beginning of the story, a flavor of how we have begun to move by thinking about principles. But let me flag something crucial about the notion of a principle. My use of this term is completely different, and ultimately wrong, relative to the level of abstraction of principles that people in linguistics have moved toward. In the case of morality, this is merely a starting point. The intuition is that as we move deeper into this problem, the abstractness of the problem will surface, and the relationship between actions, intentions, and consequences will be as complex and nuanced as are the relationships between the conceptual-intentional system and the syntactic operations that provide structure and, downstream, variegated meaning. So, when I say "principle", think of it in this looser sense, at least for now.

Let me describe three principles, with the first mapping to a distinction I just called upon: the *Intention Principle*. It is basically the principle that Thomas Aquinas invoked called the "doctrine of double effect": harm intended as a means to a goal is morally worse than equivalent harm foreseen as the side effect of the goal. Second, the *Action Principle*: harm caused by action is perceived as morally worse than equivalent harm caused by omission; and lastly, the *Contact Principle*: harm caused by physical contact is morally worse than equivalent harm caused by non-contact. To explore these principles, we developed a large set of moral dilemmas (we now have some 300-400 different moral dilemmas). For each principle, we presented a set of paired dilemmas that only differed in terms of the crucial psychological dimension captured by the principle. Subjects provided judgments on a scale from 1-7, with 1 mapping to forbidden, 4 to permissible, and 7 to obligatory. For paired cases in which subjects noticed a difference, we both evaluated this difference statistically and also asked subjects to justify their responses.

Results showed that for both the Intention and Action principles, 6 out of 6 scenario pairings revealed support for the operative force of the principle, whereas for 5 out of 6 scenarios in the Contact principle, this was also the case. Thus, subjects judged intended harms as morally worse than foreseen harms, actions as worse than omissions, and contact harm as worse than non-contact harm. The next critical question, very much analogous to questions in linguistics, was: are these principles not only operative in that they influence people's judgments, but can they be expressed? Are they recoverable, are they used consciously in deliberations of creating these moral judgments?

For the Action principle, subjects recovered a sufficient justification 80% of the time, appealing to comments such as “actions are worse.” For the Contact principle, subjects appealed to contactful harm as worse than non-contact about 50-60% of the time. Quite consistently, however, they denied the moral relevance of contact, saying such things as “Well, if you physically touch somebody and it hurts them, that is worse than if you don’t touch them... Nah, that can’t be relevant.” So they rejected the principle, and often invented assumptions to explain what was driving their judgment. But perhaps most interesting of all, very few people recovered the Intention principle. People who did failed would say things such as “I don’t know” or our favorite, “Shit happens!” So here we have a distinction between principles that in each case are clearly operative in that they are driving the nature of the judgment, but only in some cases are they recoverable in that they seem to be expressed in people’s distinctions between Case 1 and Case 2. That suggests that some principles seem to be having effects as intuitions and that maybe these intuitions are absolutely not recoverable or are inaccessible, in the same way that linguistic principles that have been discussed here in this conference are inaccessible.

One of the big questions, then, coming back to some of the themes in the Conference, is the extent to which we see these kinds of principles are universally in play. To begin addressing this question, we can pinpoint different variables that have classically been invoked as causally relevant to cross-cultural variation and explore the extent to which they influence the patterns of judgment. One of our first stabs has been in terms of religious background. As a first cut, we simply contrasted all subjects indicating some kind of religious background with those marking “atheists”. For this initial analysis, we didn’t concern ourselves with the specific kind of religious background but rather, with its presence or absence. The clear result thus far is: No. There was not a shred of evidence that people who claim to be religious showed different patterns of moral judgment or moral justification (although we did of course see people who were religious invoking more, “Well, God must have done something”), but besides that, no differences in the pattern of moral judgments. Furthermore, we found no differences between people who expressed different degrees of faith or religiosity: individuals who said that they were not very religious showed the same patterns of judgment as those who said that they were very religious; and within the limited sample that we collected, there were no consistent differences among the types of religions.

Let me digress for a moment to relate this finding to a recent experience I had in a class at Harvard, and in particular, during the presentation of this material. During my presentation, I could see that the students were getting extremely anxious. I therefore stopped the lecture and said “You all seem a bit antsy. If you have concerns or questions, please pipe up and let me know.” Upon finishing the last syllable of my sentence, one student exploded and said, “Look, I know where you are going with this. This is one of those biological, Darwinian explanations, but there is a clear alternative explanation: simply, God created all the universals.” [laughter] These are tough moments for a teacher. On the one hand, you want to respect the variety of views that people can have, and on the other hand, you want to explicate the positions, and show that issues of faith and science are entirely different ways of knowing or understanding. I responded: “We may be at an impasse here. I can either capitulate because I can’t call up any evidence to show that your position is wrong, or we can take the following path together. If it is true that most, if not all religions take as inspiration some divine power, and divine power provides the intuitions that create religious doctrine, then I think you

have a problem. Since religious doctrine can't explain the pattern of judgments we observe, but you want to argue that God or some divine power provides the universals, then you have to say that religion rejects divine inspiration when it comes to these moral judgments. This just strikes me as very problematic for the religious position, at least if you think that there is an empirical issue, as opposed to an issue that relies on faith. The other point I would make is that of course everyone taking the moral sense test logs on to the Internet, and thus our sample is very skewed. In fact you could say, "Even if you are not religious, you've been exposed to Christianity at some level, so of course that is why you are finding the pattern that you have". To address this problem, we have begun to present the same kinds of moral dilemmas to small-scale hunter-gatherer societies that have no explicit religious system — this doesn't mean that they lack beliefs, but rather, that their system of beliefs is not made explicit in the form of religious doctrine or accounts. And they certainly haven't been exposed to Christianity. So if we find similarities I think it argues even more strongly for certain patterns of universality driven by some biologically set-up system. An interesting example comes from the Kuna Indians, a very small-scale hunter-gatherer/subsistence society in Panama that has had little contact with the outside world. We have given one village community various kinds of moral dilemmas, cases that in important ways mimic the trolley problems. Now here is an intriguing, albeit preliminary result. If you give them an example that is like the bystander case, a canoe going down a river which can displace crocodiles away from 5 onto 1, 97% of the 50 or so people we tested said that it is permissible for the bystander to redirect the crocodiles. If you give them a version of the fat-man case, in which a person can push a fat man out of tree in front of a herd of stampeding boar, saving the five, but killing the fat man, 45% say that pushing is permissible. Now, here is the important lesson, I think. This culture, as well as others that we have been able to look at such as a Mayan community in the Chiapas area of Mexico, see the difference between intended and foreseen harms. In the case of the Kuna, however, the difference between intended and foreseen harms seems to be less than it is in the Western, and developed societies that we have tested on the internet. This is very preliminary and could be driven by all sorts of confounds, but for the moment, let us assume the pattern is real. We can explain the increased permissibility of intended harms in the Kuna by looking at their recent history of infanticide. That is, one sees almost no physical deformities in this society, primarily because those with such deformities are killed early in life. So intended killing is part of the society. What we think is happening, perhaps as a form of parametric variation, is that all societies will show the intention principle, but each society can tune the degree of difference between intended and foreseen harms, but not eliminate it.

I hope this provides a flavor of the argument and the work that lies ahead. I think the principles here are nowhere near where they need to be. I think in some sense we need to go back to some of the work that was started a long time ago in the philosophy of action, laying out in greater detail the nature of computation in action perception that may provide some of the primitives to our moral judgments. Even with all the empirical holes, however, I think we now have a new and important set of questions, with answers forthcoming.

Music

In this final empirical section, I will focus on music. Again, there is limitless variation in music as there appears to be for language and morality; the question is whether there are some primitives that are both driving and constraining the variation. What I don't want to spend too much time debating is a definition of music, as this could lead us down a never-ending path that would be quite fruitless. Here, however, is a quote that I like because it captures at least some of the functionality of music: "The purpose of music is to sober and quiet the mind, thus making us susceptible to divine influences." What I actually like about this quote is that it begins to capture two important aspects of music that I want to explore here, which is the interface between some kind of perceptual pattern recognition and our emotions. John Cage, of course, who really started the minimalist move in music, made exactly this kind of argument. Here is one of my favorite quotes by Cage: "I can't understand people who say I am frightened of new ideas, I am frightened of the old ones." This certainly seems to capture a flavor of the minimalist program in linguistics. There is a little piece from John Cage that relies on only three notes, over and over again, but with crucial changes in tempo and the intensity or attack. This tradition continued, for some people to their great horror, including Cage's famous "4'33" — a piece of simple silence, lasting for four minutes and 33 seconds, precisely the length of time of typical "canned music." Taking liberty with the views espoused by the minimalist movement, I think that minimalist music focuses on breaking up the intentionality created by music, emphasizing the silences, randomness, slowness, and tempo in particular. As such, it attempts to strip music to its core, its skeletal features, and assess how such structures mediate perception. Even if this rather loose interpretation is too loose, I think it is a wonderfully ambitious and exciting project that sits at the interface of the arts and sciences.

Now here is what I want to do to show again how comparative work can bear on questions of music structure and perception. It is true that every single culture that we know about has music as part of its system, and the question is, are there invariants? Two invariants that appear to emerge, cross-culturally, are that consonant intervals are perceived as more pleasant than dissonant intervals, and that lullabies have virtually identical structures, simple, repetitive elements, slow tempos, and a restricted range of frequencies. Assuming these are invariants, part of our species biological endowment, we can next ask: how did these perceptual-emotional biases evolve and are they uniquely human?

To address this question, we turn to studies of nonhuman primates. In particular, my students and I wanted to understand not only what primates perceive, but whether they spontaneously discriminate certain musical styles, and especially, like some more than others. Thus, our goal was to explore how potentially ancient perceptual mechanisms interface with the emotions to generate distinctive musical preferences.

To explore this problem, my recent graduate student Josh McDermott worked with me to design a very simple experimental approach using a V-shaped maze. We released an animal, either a cotton-top tamarin or a common marmoset, into this maze, and while they were on one branch of the V-maze, a hidden speaker played a particular sound; as soon as they crossed over to the other branch, a different sound played. What this method provides is a kind of listening station where the animal gets to choose its musical selections, at least within the options of a session. They don't receive any physical rewards for choosing, simply the exposure to different sounds. Before we explored some of the more interesting musical contrasts, we first wanted to establish

that the method would work and thus contrasted loud white noise with soft white noise. We found consistent preferences in both tamarins and marmosets: a strong preference to spend time on the soft white-noise side, as opposed to the loud. Similarly, if you present tamarins with a choice between their own, species-specific food chirps (associated with food) and their submissive screams (associated with fear), they spend more time on the chirp side than the scream side. These results reveal that the method works, providing a tool to explore spontaneous preferences for particular sounds.

Now we ask the question, if you give them dissonant intervals vs. consonant intervals, do they show a preference for consonance as would be predicted from studies of human adults and infants? Results for tamarins and marmosets failed to reveal a statistically significant preference for consonance over dissonance either at the group or individual level. This shows that neither tamarins nor marmosets show a spontaneous preference for consonance. It is unlikely that this result is due to a psychophysical constraint as several prior studies, using both behavioral and neurophysiological preparations, have revealed clear evidence of discrimination. Rather, what our studies show is that despite a physiological capacity to discriminate consonance from dissonance, this is not a meaningful distinction for these animals in that it fails to generate spontaneous preferences for one stimulus over the other.

What about lullabies? We wondered whether there would be a preference for lullabies versus something else, so we started simply, contrasting a non-vocal, flute lullaby with a non-vocal piece of German techno. The younger members of my lab were voting strongly for the techno, and I was praying secretly (though I am not religious) for the lullabies. Consistently, both species preferred the lullabies to the techno. For some, this will either represent an exceedingly trivial result or one not worth discussing because the experiment is so poorly controlled. That is, there are dozens of differences between lullabies and techno, and so the crucial question is: what acoustic properties underlie the preference for what we are describing as a lullaby? As we planned all along, the lullaby-techno contrast was simply an opening card, designed to see if we could find a systematic preference and if so, then attack the problem to determine what features are in play. I won't present all of the conditions, but will focus on one that is quite telling, specifically, the role of tempo. Recall that I mentioned the observation that lullabies tend to have slow tempos. As a result, perhaps the preference for lullabies merely reflects a preference for slow tempos. Thus, we presented a choice between short segments of sound played at either a fast (400 beats per minute) or slow (60 beats per minute) tempo. Consistently, subjects preferred the side playing the slow over the fast tempo. Thus, one crucial factor driving the preference for lullabies may be an evolutionarily ancient bias towards slow tempos. And one good reason for this preference is that if you look at a whole variety of species' alarm calls, they are typically associated with fast tempos. Fast tempoed sounds seem to be coupled with aversion or avoidance of what is going on in their natural vocalizations.

Following a discussion of this work, several colleagues challenged us with different versions of the following question: "Okay, your tamarins might prefer lullabies over techno, and slow over fast tempos, but do they prefer lullabies as we seem to do, and as children do, over peace and quiet?" The answer is a resounding "No!" Both species actually prefer silence to hearing a lullaby. So even though they have a preference for certain kinds of music, there seems to be a strong preference for silence over noise.

Perhaps they are just ahead of their time, prescient animals who had to wait for minimalism and John Cage's "4'33".

Coming back to some of the themes of today, what might be the uniquely human aspect of the music faculty is the interface between evolutionarily shared systems of tempo and frequency discrimination together with the systems that are recruited for emotional processing. That is, we share most, if not all of our capacities for frequency and tempo discrimination with other animals, and a significant proportion of our abilities for emotional processing with animals, but it is the interface between these systems that perhaps uniquely constructed our music faculty.

Two final points to wrap up. What I have tried to argue in this talk is that one way of thinking about the nature of the human mind is to take the lead from much of what is happening in biology much more generally, what's been happening in linguistics more specifically, and running with the idea of universal minimalism, the idea being to look for basic rules and computations. I think this is consistent with some of the issues that Chris Cherniak brought up this morning. If you look at some of the core computations that have been invoked in this Conference and for the minimalist program more generally — notions like Copy, Move, Merge, Hierarchical Dominance and so forth — these are precisely the kinds of operations that are invoked by cellular and molecular biologists such as Mark Kirschner. Secondly, once you invoke notions of modularity such as those that Gabby Dover brought up this morning, you somehow need to create mechanisms that will enable interfaces between systems. The crucial question is: what is doing the translation, and how do the different representational formats "speak" to each other? In the case of language, for example, how does the representational format that codes for distinctive features in phonology interface with the representational format that codes for concepts within the system of semantics? Lastly, given the promiscuity of these systems to create the variation, ultimately what happens is that the environment is going to prune them back from the biologically-given options, and this process will yield the distinctive signature observed in the local environment — thus, the move from I-language to E-language.

I hope this gives a reasonable sketch of the minimalist approach, and how it might open the door to new ways of thinking about our minds and how they evolved.

Thank you.

END

MARC HAUSER: We are open for questions.

JAMES HIGGINBOTHAM: I wanted to raise a question on intended vs. foreseen. I think it is a bit tricky to make the distinction. You may remember that Kant famously said that you intend the consequences of everything you intend. So in the sense of Kant's dictum, in using the weight to save the 5 people, I also know that death is a consequence and therefore I intend the death of the skinny guy. Moreover, it is a bit of a trick if you ask how these things are conceptualized. Suppose I intend to take a drink of water. So then I stand up and I walk over and I pick up the bottle. Or I flip a switch because I want to find my eyeglasses. If you ask what I intended to do, one might view

the situation in the following way. I intended to move my finger like this [stretching finger out] and the rest [moving arm forward] was foreseen consequences. So I think you have to frame it in some way that is rather careful, where you speak of the intended-foreseen distinction in some way that is categorized properly for the agent, and it is not so obvious how to do this just from a description. Also a correlative question is, I have read a number of books, all of which I think are terrible, about moral permissibility, as if this were some kind of abstract stuff that you can sort of throw out, right? But is it possible to change any results, or have you considered asking the question in a more first-personal way? Would *you* pull the trigger, vs. should *he* pull the trigger — the dirty-hands thing?

M. H.: Yes, great questions, and I am sympathetic particularly to the first. I mean I think I was trying to foreshadow your question in the sense of saying that I think the notion of principles that I am picking up is really crude, and I think it is crude in precisely the way that you pointed out. For example, try pushing the analogy to language even further. Let's say that the notion of an action, a representation of an action, or what we might call an "acteme", is like a phoneme — completely meaningless in isolation, and only gaining in meaning as a function of particular sequences, underpinned by intentional states, and generating particular consequences. In this sense, bending your finger may be either meaningful or meaningless. It depends on how it is strung together with intentional states and other surrounding actemes or actions. John Mikhail, who has written a very nice thesis and is really my co-collaborator in much of this, intellectually at least, has tried to make much more subtle kinds of distinctions appealing back to some of the philosophy of action, and especially Goldman-esque decision trees. I think the problem is that these trees are not at the right level of grain. And I think all the complications that have been raised in the philosophy of mind and language about the notion of intentionality are not cashed out. Frances Kamm, I think, is one of the few people actually engaged in moving these ideas forward. She has created extraordinarily complex dilemmas that largely target the same scenario, the famous trolley problems. For Kamm, however, the issue is not one about empirical research or deep questions about the mind, but about probing our intuitions to decide what is prescriptively or normatively permissible. That said, I am convinced that the kind of work she has put into play will make significant contributions to the empirical studies that we are engaged with.

On the second point, we have approached this question from several other directions. Let me tell you about two of these. If everyone carries around some version of the categorical imperative, then they should answer these scenarios in the same way if they are judging a) their own actions as the bystander, b) a third party as the bystander, and c) themselves or another as one of the possible victims on the track. This would be exquisite evidence of a folk theory of the categorical imperative! Now the problem is, how to dissociate what will clearly be a very strong emotional response to being on the track and saving your derriere. And this is precisely where studies of patient populations enter, and in particular, patients with damage to brain areas associated with emotional processing. This isn't going to directly answer the You-I suggestion, but let me give you a flavor of the move.. Consider Tony Damasio's studies of patients with damage to the ventromedial prefrontal cortex patients. Much of the work on these patients suggests that there is a problem with the connection between the emotional areas in the amygdala and decision areas (this is crudely described) in the frontal lobes. Due to such deficits, these patients appear to have severe problems in the socioemotional domain, including

moral decision making. In collaboration with Damasio and several other colleagues, including my two students Liane Young and Fiery Cushman, we have refined our understanding of this deficit by systematically exploring a broader part of the space of moral judgment. Cutting a long story short, these patients seem to have a very selective deficit: they only show differences between normals on a certain class of dilemmas which are other-serving personal cases. They are true moral dilemmas in the sense that there is no adjudicating norm that clearly arbitrates between the options, and where one option is to engage an action that is aversive, but where the consequence is to maximize aggregate welfare in terms of saving more lives. Under these conditions, the frontal patients go with the utilitarian outcome, as if the aversiveness of the act was irrelevant.

J. H.: You said the categorical imperative, but actually you meant the utilitarian, I think. It is the utilitarian for whom it doesn't matter who carries out the action.

M. H.: I meant the categorical imperative in the sense of, I think this is a permissible action, I just think it is permissible in the sense of being permissible for any...

J. H.: An imperative is an imperative about maxims, it is not about individual actions, it is about reasons for doing them. It is the utilitarian who has the problem here.

M. H.: Right. More questions?

1:04:03 [U. Barcelona professor] CARME PICALLO: If I didn't misunderstand you, you related the lack of lexicon in primates to an inability for vocal imitation. Is that right?

M. H.: That is not the sole reason. What I was saying is, that no matter how rich the conceptual system, there are at least two problems, one of which is that even if they could externalize, they can't pass the information on. So there are two problems. Thinking about FLN again, there is both the problem of the mapping between sound and meaning, but there is the additional problem of being able to pass it on.

C.P.: Yes, but then what I am questioning is your saying it just in terms of vocal imitation, because certainly they can pass on gestures, but they cannot pass on signs, or sign language.

M. H.: Even gestural imitation is extremely impoverished in primates. "Monkey see, monkey do," just for the record, is a myth. No evidence. The best gestural imitation is weak, very very weak, relative to humans. It has taken literally thirty years to show even the most slight evidence of it. So it is absent vocally, it is weak at best visually.

NOAM CHOMSKY: On this same point, there is another form of transmission, namely by inheritance. So suppose that you get a smart ape, one that comes up with a combinatorial system. That ape has advantages. It can think, it can plan, it can interpret and so on, and its descendants will have the same advantages even without vocalization. If those advantages are sufficient, they could take over the whole breeding group. They'd all have these capacities and then vocalization could come along later because it is useful to interact. So I think there is a crucial — I'd like to expand the difference between Steve and me — I think a possibility is that that is the way the transmission took place.

M. H.: Yes. My quick answer to that is that that would not affect the story I told. Indeed, it would add to it, which I think is perfectly reasonable. In fact, it does enlarge the gap because it says that much of primate thought could have been really moving in

quite extraordinary ways by genetic transmission=, and then it may have even been a more simple trick of something about the auditory-production loop that got fused, and that could have been a trivial step.

N. C.: Yes. The other question is whether you are proving something that I have always believed, namely that teenagers are a different species [laughter].

M. H.: Like right, man.

MASSIMO PIATTELLI-PALMARINI: Marc, I was wondering whether you have in the domain of decision-making..., as you know Thomas Gilovich and Kahneman and others have shown that there is more regret for something you did than for something you didn't do, even though the consequences are exactly the same. And this is a traditional thing, and you seem to have it here, you know, omission vs. action. But on the other hand, other data from **Connally and Zellenberg** and others have shown that unless you are somebody who is supposed to be doing something — for example a doctor is called to do something [and he either does nothing, or he does something and it was wrong]. The same side-effect happens when the doctor did nothing, versus the doctor did something and it was wrong. The doctor that was called to do something and did nothing is considered worse morally. So I wonder whether you plan to extend it. It would be interesting because it has to do with counterfactuals, it also has to do with a number of other tests that have been made in which, you know, where there is something anomalous in the series of actions, you pinpoint the anomalous thing as being *the* cause of what happens.

M. H.: Yes, we are. Take the classic case that James Rachels brought forward, which some of you may well know, of the greedy uncle who wants to do away with his nephew who is first in line for the family inheritance. In Story 1, he's babysitting the nephew and he goes upstairs while the nephew is taking a bath, and he intends to kill the kid and he drowns him. So he intends to kill and he kills. In Story 2 he has the same exact intentions, he goes upstairs, the kid has flipped over in the bathtub and is drowning, and he just walks away and lets him drown. Now the first is an action, the second is an omission, but we don't want to see those cases as different, in fact we see them both as the same. So in some cases we don't see a difference between action and omission, and in some cases we do, and the question is, how much information do you attribute to the agent, that then makes you either lose it or pick it up? I think it is not clear in the philosophical literature at all. And it is also not clear to what extent we are vulnerable to the action-omission distinction. So that is what we are trying to do in two ways. One is to play around with when you get the information, whether you get the consequences first or the intentionality first, and I will come back to that in a second, and the second question is, to what extent is this distinction available early in life? We know almost nothing about action-omission in its ontogeny in young children; we know almost nothing about the intended-foreseen distinction either. Oddly, even though there has been a rich literature on theory of mind, these distinctions don't enter into the discourse because they have largely been developed within moral psychology and the law. But they bear directly on the agent's mental states.

One more case. This is a nice case by a philosopher named Joshua Knobe. There is a CEO of a company, and the President of the company comes to the CEO and says, "Look, I've got a policy which, if implemented, will make us millions of dollars". Now there are two versions of the story. In Story 1, the President says, "If we implement the policy, there is a good chance it'll make millions of dollars, but it will probably harm

the environment”. In Story 2, the policy will probably help the environment. In the first case, The CEO says “Look, I don’t care about harming the environment, I just care about making money. Implement the policy.” The company implements the policy, and they make millions of dollars, and it harms the environment. Now you ask people: did the CEO intend to harm the environment? Here, subjects say yes. In the help case, however, they say no. The idea here is that labeling an individual with some kind of moral attribute like *blame*, or *blameworthy*, actually affects the intentional attribution, at least at some level. So again, there can be all sorts of ways in which these patterns unfold, depending upon the temporal flow of them through time.

JANET FODOR: Making these moral distinctions is actually very distressing. I have to kill at least one guy if not 5. And so the natural thing is to reach for a rationale, an excuse of some kind, and a very common excuse it seems to me is to blame the victim. I think we can all remember cases where we have tried to blame the victim. So this is a factor that could be introduced into the experiment so that the 5 guys have been told it is stupid to walk on the tracks, it is dangerous to them and everybody else; the one fat guy has been told that this is a track that is never used, it is perfectly safe to walk here, or vice-versa. Is this a universal that would make a difference to the study?

M. H.: Yes, the problem here is that the space at some level is unconstrained in terms of the number and kind of permutations you could run. You could ask about in-group vs. out-group, you could change the numbers — there are all sorts of things you could change, and several papers have explored this part of the problem. We have taken a different route, which is to hold these personality traits constant in order to explore the causal-intentional structure of event perception in the moral domain. As soon as you put things like responsibility (like they’re workmen, they should be there; or it is the conductor’s job, he’s got a responsibility towards the others), you are going to change lots of the dynamics. And I find these to be very difficult problems, headed more toward social psychology, and a zone that I am less familiar with. It is not that these are the wrong kinds of questions, but rather, that I have less confidence with regard to the experimental questions. . I feel more comfortable with the primitives underlying causal-intentional attributions because I am quite convinced that we can explore these issues in infants, animals, patient populations and so forth. That is, in the same way that Lila Gleitman was exploring the foundations of giving and hugging in infants, looking at how infants dig beneath the surface dynamics, we have moved in a similar direction.

What is also of interest is how certain people can be about their judgments for one or two cases, but as they pile up, they lose this confidence, and the reason we think this is happening is because they don’t really have access to the underlying principles, just the surface features of each case. Let me illustrate with my father, a very smart, rational physicist. He had asked me what I was working on, and so I decided to illustrate by giving him some moral dilemmas. I started with the bystander case and he answered, “Well of course you flip the switch”. I then gave him the fat man trolley case and he said, “Well of course you push the man”. This is, you will recall, a relatively rare response so I asked him why. He replied “Well, because it is always better to save 5 than 1”. I then give him the classic organ donor case, where there is a surgeon in the hospital and a nurse comes in and says, “We’ve got 5 people in critical care, each needs an organ, we have no time to ship out for the organs, but you know what? This guy just walked in to visit his friend. We could take his organs and save 5 lives. Is that okay?”

My father says, “That’s ridiculous, you can’t just kill a healthy person off the streets!” “But wait,” I say, “you just killed the fat man.” He says, “Okay, you can’t kill the fat man.” I say, “What about the switch case where you killed the guy on the side track?” He says, “Ok, you can’t do that either.” So, ultimately, the whole thing unravels, because you can only locally explain one dilemma but you can’t explain the cluster, because you don’t have access to the underlying principles. This is the core intuition driving our work.

J.F.: I am not sure why you think the universals are likely to be about the agent and not about the patient.

M. H.: Oh, I think the universals may come in at the patient level too. My guess is that it is a universal, and so there are studies by Lewis Petrinovich showing that if you put kinship as the patients, you are going to get evolutionary sociobiology to work. “I will favor those who are more genetically-related to me than those who are not, all else equal.” You can get species effects. “I will favor human over even an endangered species like the chimpanzee.” So these effects are certainly operative, and they may well be a part of what is universal, but I don’t think this part is specific to morality. Ingroup-Outgroup distinctions arise in all sorts of contexts, some moral and some not. More importantly, perhaps, we have tried to tackle a different set of problems by holding patient identity out of the scenario, operating under a kind of Rawlsian veil of ignorance. By doing this, we hope to uncover the architecture of the underlying psychological cause of the agent’s actual action.

PARTICIPANT: I wanted to shift to the musical section. I’d be very curious, if you could create a variable to your experiment — I would suspect that part of the preference for silence has to do with just the foreign nature of technologically-produced sounds that we have *learned* to appreciate, like the flute or recorder sound that you had in the lullaby. It may be quite offensive to the ear of some other primate. Referring back to your earlier experiments with quantification, one of the appeals of music is the structures that with repetitious quantities — a performing musician learns to play chord progressions, for example, without physically counting them. You don’t go “One, two, three, one one, two three two”; you *hear* the changes. And I wondered if there would be an appeal among primates that had a quantifying capacity, when using sounds from nature, that could be organized structurally to repetitively use sounds that they’re familiar with rather than some kind of human technology that is used. Just an idea to look at whether that quantifying..., the appeal of the recognition of repetitive quantities, is a big factor in what we like about music. It’s why young people like Techno.

M. H.: In some sense I agree, but we are at such an early stage of this work that it is hard to make sense of much of it. There have been a couple of papers recently, by Smith and Lewicki, claiming that lots of the physiological firing patterns that you see to sounds have a very primitive system or structure that really taps into natural sound, and speech may simply be parasitic on this mechanism. This position or perspective sets up a study that I just finished in collaboration with Athena Vouloumanos) — this is a slight tangent with respect to your question but it gives you an idea of what would be basic in terms of the auditory system. There has been fifty years of research on neonates’ preference for speech. The common lore is that babies are, early on, tuned to speech, preferring to listen to speech than non-speech, and showing significant abilities for speech discrimination. The problem has been that none of the studies to date have contrasted speech with other biological sounds, focusing in stead on reversed speech,

sign wave speech, white noise, and a variety of non-biological sounds. . We therefore set out to test 32 neonates, less than 48 hours old, with a non-nutritive sucking technique where suck rate gives information about interest or attention to the material. We contrasted non-native speech with rhesus monkey vocalizations, and thus, both sets of stimuli are novel at some level. The result was clear as can be: no preferences at all. Neonates sucked as much for non-native speech as they did for rhesus monkey calls. By the age of 3 months, however, possibly earlier, a preference for speech is in place. These results suggest that there general auditory biases that get tuned up quickly in development. These results also rule out all the arguments that have been put forward for *in utero* experience, because by the third trimester, the baby is certainly getting some acoustic input. But whatever that is, it is insufficient to create a preference for speech over rhesus monkey calls. So that is just a long-winded way of saying that some of the preferences in music may well derive from quite general auditory preferences. It is certainly possible that if we had played more biological sounds, perhaps structured in some musically relevant pattern, that we would have seen a different pattern of responses. Let me add one relatively new piece of data, still preliminary in terms of our analyses. We have just completed a study in which we presented marmosets with 5 months of exposure, 12 hours a day, to consonant chords, and then, to samples of The idea here was to more closely approximate the kind of exposure that human infants receive during early development. When we subsequently tested our subjects for a preference for consonance over dissonance, either chords or pieces of Mozart, subjects showed no preference. Interestingly, however, infants exposed to the same materials showed a mere exposure effect, preferring the specific sounds played over novel, but matched sounds. This provides one of the first pieces of evidence in a nonhuman primate for a critical period effect. **ROCHEL GELMAN:** Just a bit on music. I expected you to be talking about something like harmonic principles, some principles of music that the mind treats as privileged. Consonance-dissonance is one of those, but if you are using chords, it matters whether you are in Western music or not. So the issue becomes what's universal across different harmonies. I believe it is the case that the octave and the fifth appear almost invariably in every harmony, where it is consonant, and that the transitions are such that they'll be very different, but you will find they are the fifth. That is not trivial because the physics is such that the first overtone's the octave, the second is the fifth, the ear is sensitive to these, etc. So this might mean that the principles are highly abstract and are harmonic principles, just as the linguistic principles that we are looking for are very highly abstract. And maybe it is not a question of whether the sounds are consonant are not, because what's consonant for us is not necessarily consonant in another culture. It is how we fill in the transitions that varies enormously.

M.H.: Yes, it is a crude cut. And again, psychophysically there have been studies showing that animals exhibit octave discrimination and generalization. In particular, Anthony Wright has shown that if you train rhesus monkeys in a match-to-sample paradigm, using children's melodies, they can readily do the transpositions.

R.G.: The really interesting question is whether they will also generalize to the fifth; actually I have data I never published that shows 5-year olds will. But they're experienced.

M.H.: Way experienced.

Thank you very much.