# CHAPTER ELEVEN
# MEMES, MIND AND NORMATIVITY

YUJIAN ZHENG

> Darwin's theory of evolution …is counter-intuitive and hard to grasp but once you have seen it the world is transformed before your eyes. There is no longer any need for a grand designer to explain all the complexity of the living world. There is just a stark and mindless procedure by which we have all come about – beautiful but scary.
>
> Susan Blackmore

**ABSTRACT** Prominent memeticists like Daniel Dennett and Susan Blackmore have made claims far more radical than those included in Dawkins' original proposal, which provoked increasingly heated debates and arguments over the theoretical significance as well as limits or flaws of the entire memetic enterprise. In this paper, I examine closely some of the critical points taken by Kate Distin in her penetrating engagement with those radical claims, which include such ideas as the thought that we are meme machines as much as gene machines, the thesis that there is no conscious self inside those machines, and the claim that a complex interplay of replicators and environment is all there is to life (Blackmore 1999: 241). It is hoped that a viable thesis concerning a deep-seated normativity emerges from my discussion.

## 1. From Genes to Memes

To a scientifically informed reader, the above remark doesn't sound exaggerated. Yet there is some ambiguity in the last sentence, i.e., how much of our own mindedness, with its splendid products (all subsumable under "culture"), come about by that same "stark and mindless procedure?" Richard Dawkins first proposed his version of cultural evolutionary theory in The Selfish Gene (1976), whose main thrust, though, was a defense of the gene as the unit of biological selection and the organism as a "survival machine" for its genes. The obvious explanatory power of the genetic hypothesis (even long before the discovery of its physical basis in DNA) is not a topic for discussion here.

What is interesting is an implicit twist near the end of Dawkins' book that led to his coinage of the term "meme" as the unit of cultural selection, which has not only been added to the Oxford English Dictionary, but developed into a growing domain of studies called "memetics." The twist is this: ideas and customs in culture develop at a pace that is far too great to be picked up at the level of biological evolution (just think of the

headlong rush of computer-related technology). Sociobiology, as an ambitious predecessor of memetics, attempted to show how the evolution of the body could account for cultural changes and, unsurprisingly, was doomed to failure due to this very disparity of pace between the two kinds of evolution. Unlike E. Wilson who believes in genetic fitness of every successful aspect of culture, Dawkins suggests that perhaps we should look instead to the analogues of genes within culture itself, i.e., memes, which would provide a mechanism for cultural evolution. On this suggestion, biological Darwinism becomes an example of a general type of theory, called Universal Darwinism, whose three key features are: variation, selection, and replication (or heredity). In other words, any physical or ontological item can serve as replicators or units of selection as long as their environments have limited resources or exert pressures of differential survival, and certain conditions on variation and retention are met.

## 2. Content and Norm

The first question for memetics is naturally "what are memes?" Dawkins' examples include ideas, catch-phrases, tunes, gestures, fashions and skills. Though intuitively plausible as candidates for replication, these things are of quite different natures. What seems unclear about such a list is whether there are any unifying or underlying conditions that constitute or guarantee their status of replicators. For instance, Distin suggests that memes must have enough stability to be particulate, which means that "they must be able to slot into established cultural assemblies without their own informational content being lost or blended in the process" (Distin 2005: 198). Yet at the same time they must not be too rigid to mesh well with other components in a variety of cultural contexts. A related issue is therefore the criteria for the identifiability of these supposed units of replication. For example, should we take the smallest meaningful parts of a speech or thought (say, a phoneme) as a meme, or, for the sake of the integrity of content, take certain large enough constructs of a representational system as the authentic memes?

Common to many such issues is the very notion of representational

content, with its relationship to the phenotypic effects it is supposed to be about and/or control. When a person walks to a vendor machine with some coins in hand, a true ascription of his or her mental states may involve the belief that putting a certain number of coins into the machine will cause it to yield the desired snack. Such a belief, in conjunction with certain other relevant mental states (e.g., a desire to eat the snack), will suffice to explain the observable behavior, which is a phenotypic effect.

The representational content of this belief is not merely about those particular external objects such as the coin, snack and vendor machine, but also about the particular means-end relation obtainable between these objects. Representation of this latter relation must figure in whatever control mechanism underlies the behavior. More pertinently, the use of such representational content is not limited to this particular context, but will extend to similar or different contexts involving that means-end relation. The relevance to memetics of the context-transcending character of representational contents is the point that they are replicable interpersonally as well as intra-personally. The prime reason for regarding beliefs or ideas as (potential) memes lies exactly in the fact the fundamental nature of conceptual contents allows them to be learnable or reusable across individuals and thus liable to widespread dissemination, especially when their utility proves great.

Compare the above example with those involving animals. A monkey may be able to use a stick to get a banana hanging up from the ceiling. Such behavior may be repeatable when similar contexts reappear. The external means-end relation between the stick and the banana seems to be somehow utilized or internalized by the cognitive state of the monkey. The question, however, is under what condition, and to what extent, the representational content of the monkey's internal state can be, or ought to be regarded as definite and distinct enough to count as meme or replicator in some analogous sense. A crucial aspect of this question concerns the independent status of any memetic content vis-à-vis its perceptual input (the external causal object) and motor-control output (the behavioral phenotypic effect), rather than about how easily or readily such banana-getting behavior is learnable by other monkeys, though the

latter issue is also relevant. In the case of monkeys, let alone other lower animals, their internal cognitive states could hardly have any memetic content independent of those causal inputs/outputs. Yet how can one be so sure about this? Shouldn't it be an empirical issue that is better left to science to settle?

Although many contemporary philosophers have much to offer on this issue, I'd like to confine my discussion to c who, in The Selfish Meme, presents a quite straightforward line of thought which shares its main tenet with a mainstream holistic view in epistemology and semantics. The gist of the thought is this. For any candidate for being a unit of replication, whether it be at a biological or cultural level, the important point is that real replicators exist independently of their effects. It means that even when their phenotypic effects appear to blend or don't properly manifest at all, we should have a method to identify or individuate the contents of these replicators. In the case of cultural replicators or memes, their contents must first of all be representational.

But representation itself (at least if we use the term "representation" in a broad, inclusive sense) is not sufficient for qualifying its content as memetic. Why? Because memes are supposed to preserve their determinate content in a way that can be replicated between individuals under wide and diverse contexts (including across distant spaces and generations). In order to account for the breadth and depth of human culture, memes must be able to interact and assemble with other memes in a way of forming internal links among various representations, in addition to their links with external perceptions and behaviors. "[T]hese internal links give representations their internal properties such as identity, and ultimately free them from their dependence on external stimulation" (ibid. 200). A crucial, distinctive feature emerges at this point: first-order representations can now be meta-represented, i.e., independent symbols are invented to represent relations among representations that may bear various structural similarities or differences at higher-order or more abstract levels. It is the meta-representation, claims Distin, that ultimately separates human culture from all animal "cultures" and enables human representations to gain independence from

their original contexts, developing in complexity and abstractness.

Moreover, it is also due to our ability to meta-represent that we acquire our own natural languages (which share fundamental "universal grammar" according to Chomsky's broadly accepted linguistic theory). More interestingly, Distin proposes that this same ability underlies the development of nonlinguistic representational systems (RS) such as those of mathematical and musical notation, whose diverse rules and structures are realized in media other than speech. An important implication is that many novel concepts are not available to us prior to the development of the (specialist or general) RSs that support them.

The above line of thought is largely plausible, and probably by now well belongs to the mainstream position in philosophy and other related disciplines. A remarkable, key point I'd like to emphasize here, however, is the fundamentally normative nature of the emergence of so-called meta-representation or symbolic thinking in general. Uncontroversially, symbols work "via social convention or established code, rather than by resembling or being straightforwardly correlated with that which they represent (ibid., 147-8)." What underlies the possibility of forming conventions or stipulations is clearly a normative matter, regardless of the detail of the causal story behind it. But what strikes me as puzzling and unsatisfactory, however, is Distin's insistence on the innateness of the ability to meta-represent. Her confidence of saying so seems to derive mainly from Chomsky's postulation of a language instinct, combined with the thesis that natural languages have primacy in developing all other kinds of RSs. The "language instinct" may be a good metaphor or explanatory device for making sense of certain uniform features of linguistic phenomena, but it doesn't automatically answer, even speak to, the philosophical challenge addressing the undeniable normativity issue. It is an empirically testable fact that a human infant, with the same language instinct, couldn't develop any symbolic thought by himself were he raised in a non-human environment, or heavily deprived social contexts. Thus it can be seen that normativity is a distinctive, social feature which can't be equated or reduced to natural capacities of any kind, even though the latter must be presupposed in the causal

explanation of the former.

A major difference between normativity and any genetic heredity or instinctive constraints lies in the fact that normative constraints don't have natural necessity in the sense that the subjects cannot violate it under any circumstance. Rather, human rules and conventions are self-imposed (at least in a collective sense) and subject to self-conscious violation or defection as well as critical revision or abandonment. If all memes carry normative content (or component), as seems clear in the above depiction of Distin's views, then it is unsurprising that the greatest disanalogy between meme and gene should manifest itself in the following difference: gene tokens are causally prior to, and in fact produce, the organism's body which is necessary for preserving and transmitting these very genes; whereas meme tokens are causally dependent upon, and sometimes created by, the individual's mental function/process which is necessary for preserving and transmitting these very memes. In other words, the genetic message is built in the gene tokens while the memetic message is largely external to the meme tokens (e.g., word forms or phonemes), saliently depending on the larger RS or "web of beliefs" in which they are located – that presumably explain the necessity, as well as difficulty, of interpretation for any cultural communication. And obviously interpretation is a rule-guided or norm-constituted activity, sensitive to specific contexts.

From such a disanalogy, one seems entitled to conclude that ontologically speaking, the meme is not entirely independent from, or parallel to the physical existence of gene, but is rather supervenient on the latter (plus some other special environmental conditions). This is not to deny, however, that the notion of "meme" is methodologically useful, or even necessary, in our theoretical attempts better to understand complex cultural phenomena. Moreover, it may be conceptually compatible with the above conclusion to claim that memes are real in their psychological effects on us as real persons and in the relative stability or integrity of their content. To assess such a claim, we need turn to another big issue raised by memetics.

## 3. Rational Pattern-Recognition and Deep (Anthropic) Normativity

As Richard Dawkins remarks, the fundamental insight of his selfish gene theory is "that there are two ways of looking at natural selection, the gene's angle and that of the individual" (Dawkins 1989: viii). And these are "two views of the same truth (ibid., ix)." Correspondingly, the essence of any selfish meme hypothesis derivable from it must be that there are two ways of looking at cultural change, the meme's angle and that of the human mind. The ultimate challenge here, however, is to make the case that the consciousness (or freedom, or intentionality) of the mind's viewpoint is compatible with the unconsciousness (or autonomy) of memes' mechanisms of replication or evolutionary algorithm. Again, let me first follow Distin a little bit in her catchy exposition of the problematics.

Memes don't "build" human minds in the same way that genes build our bodies as the vehicles for their own propagation and protection. Yet we may still say that a human mind is partly the product of the memes that bombard it, in the sense that the mental faculty could not fully develop and play its vital role in cultural evolution without the acquisition of existing memes. On the other hand, however, such a bombardment occurs only because the genetically rooted mental faculty has the innate potential to interact with and develop in response to those existing memes. In short, "minds are the unique product of an interaction between two quite independent Darwinian processes, one biological and the other cultural. The first is responsible for the mind's innate potential, and the other for the realization of that potential" (Distin 2005: 203).

Distin believes that such a view can reconcile the two seemingly incompatible perspectives, i.e., the evolutionary algorithm vs. the perspective of the individual mind. Minds do have consciousness and genuinely choose, and it is exactly such active mental processes that provide the distinctive mechanism of replication, variation, and selection of memes. "The consciousness that is involved (at least some of the time) in these mechanisms serves no more to undermine the unconsciousness of the cultural evolutionary algorithm than the emotions and awareness that are involved in human sexual reproduction serve to undermine the unconsciousness of the biological evolutionary algorithm" (ibid., 204).

The prima facie plausibility this view may possess doesn't, however, eliminate some deep-seated perplexity about the profound implication of taking an evolutionary theory of culture seriously. The overarching controversy is this. On the very last page of The Selfish Gene, Dawkins makes an upbeat call: "We are built as gene machines and cultured as meme machines, but we have the power to turn against our creators. We, alone on earth, can rebel against the tyranny of the selfish replicators."

Nick Rose, for one, accuses Dawkins here of having "committed a common, but fundamental error; he assumed there was 'someone' beyond the constructs of the memes and genes who could do the overthrowing" (Rose 1998: 5). The target here is the commonsensical idea that one's conscious Self can select or design the very memes that comprise one's mind, which Rose finds contradictory with the memetic assumption. A related idea is the so-called "directed mutation" – another paradoxical position, for Rose, whereby the memetic mutation that occurs is not random, but is somehow directed by human intentionality towards some goal. It's paradoxical because it violates, under the Darwinian guise, the very Darwinian principle of blind selection. Rose asks the question: if we can intentionally design memes, why do we need an evolutionary theory of culture at all? In other words, "if one believes that 'consciousness' has the foresight and independence to select and direct behaviors towards some goal, one need not posit evolution to explain the forms we find in culture" (ibid., 6). Dennett provides an earlier, powerful ally here:

> But if it is true that human minds are themselves to a very great degree the creation of memes, then we cannot sustain the polarity of vision with which we have started; it cannot be 'memes versus us' because earlier infestations of memes have already played a major role in determining who or what we are. The 'independent' mind struggling to protect itself from alien and dangerous memes is a myth ... (original italics) (Dennet 1995: 207).

The selective filters of memes are not the mythical Cartesian Self but rather are constructs of earlier memes and genes. This is a view that cuts deep into the "illusory" nature of autonomous consciousness, intentionality or freedom. For Dennett, memes are not seen merely as

replicators within consciousness, but as the essence of consciousness itself. We are made of selfish replicators, and they do not merely colonize our otherwise healthy minds – they are our minds, healthy or otherwise. Clearly, in this view rebellion against the tyranny of the replicators is impossible. Since everybody is the product of such replicators, "at the end of the day who would escape the tyranny of the replicators?" (Rose 1999: 2).

How would Distin or any meme-mind compatibilist (like J. Wilkins) respond to such challenges?

A main tactic available to the compatibilists is to drive a peaceful wedge between the big picture of cultural evolution and the small picture about individual centers of agency which can only exercise limited amount of control, however real, over the social selection process at large. In that way, it aims to keep intact both the unconscious, impersonal character of the direction of memetic evolution and the rational, personal character of individual choice. The crucial linkage between the two pictures seems to be that the local mechanisms of variation, replication, and selection are provided by each individual mind, whose status of being heavily outnumbered by other minds almost guarantees that its contribution, no matter how creative, rational or thoughtful, is bound to be limited. So, in the statistical sense, nobody can have the ultimate foresight or power to make his or her choice always the fittest in such a vast field of memetic competition. In other words, the fate of any idea, new or old, brilliant or hackneyed, is determined by its relative fitness in the world stock of memes. To have such an evolutionary fate, we don't have to deny that the idea depends on individual minds for its content to be understood, transmitted, or created in the first place. As John Wilkins remarks, "selection operates on outcomes not the provenance of variations" (Wilkins 1999:1).

This tactic seems to be able to save the macro evolutionary effect from the micro nonrandom variation. But does it suffice to rebut Rose's objection to the "Self-centered variation"? If ultimately the ontic Self or mind is illusory, as Dennett and Blackmore urge it is, this tactic cannot work for the compatibilists. And for those who genuinely endorse the

doctrine of memetic evolution, Rose's challenge remains: as far as understanding social selection is concerned, the role of individual intention is redundant, i.e., the utility of intentionality comes from its convenience as a short-hand for a more complex underlying process involving replicators of various sorts at various levels. A related idea behind this challenge seems to be this: if human beings, including their mindedness, ultimately come from nature with its evolutionary mechanism, and we also already subscribe to the big evolutionary picture of culture, then what space is left for insisting on the miraculous status of autonomous intentionality beyond the tyranny of replicators?

The answer, if there should be one, lies, I believe, in some profound thesis of deep normativity, which has not been clearly formulated by any theorist in the field. I'd like to attempt a quick and rough formulation of it.

The central concern here is how to understand the status of intentionality in relation to role of natural selection in its very coming to be. What is so special or exceptional about intentionality in a thoroughly naturalist universe in which "Man" is not separate from the rest of nature?

Let us think about science, arguably the greatest achievement of human intentionality. Nobody would deny that science aims at truth, or manages itself to discover laws of nature, whose objectivity is, by definition, not subject to random variations of human ideas. Although historians of science may cite numerous examples of wrong scientific theories or delineate the tortuous paths along which successful paradigms got established, the key point here is the built-in normative objective and criteria of this rational enterprise. Again, few would deny the great success with which science has so far lived up to such criteria (including the success Darwin's theory enjoys). Could memetics alone account for such a predictable "fate" of science and the scientific community's rational control over its results?

The deeper, philosophical question here is: how is truth possible? Or, what is the epistemological condition for any creature to know any real patterns in the world? Without addressing such questions, one could hardly see in what sense the issue about the status of intentionality is

resolvable. For even the evolutionary theory itself depends on the knowability of real patterns in nature for its credibility or explanatory validity. Here is a brief sketch of my thesis regarding a "deep normativity" built in the notion of intentionality.

Only when authentic (i.e., human) representation appeared or emerged from (whatever underlying mechanisms of) evolutionary process, did it ever, for the first time become the case that patterns and levels are recognizable as patterns and levels (and, accordingly, the possibility of error or mistake as misrepresentation is "opened"). If the phrase "unrecognizable patterns" seems contradictory in terms, then one had better say that the late emergence of human intentional or representational power is responsible for "making" patterns in nature which are supposed to be possessed by things appearing much earlier in the cosmological history. This marks a pivotal turning point at which everything prior to us starts to gain "new meaning," if you like. For instance, the proofreading enzyme does not recognize the error it corrects qua error – it makes this sense only after some representation (of certain relevant functions) is available. Put in another way, if one has to treat human intentionality as nothing but derived intentionality from mother nature, as Dennett does,[2] one must acknowledge that it is precisely owing to mankind, this unique type of "derived meaner," that everything else in the world has "meant" what it truly means (since some moment of evolution which might not be exactly pinpointable).

Theoretically, one could always assume a certain timeless, omniscient Interpreter's perspective (a "God's eye view" if you want) from which every real thing or event objectively possesses or partakes of real patterns, but such a perspective itself is possible only when human representation is in order. In short, the "Unmeant Meaner" (as comparable to Aristotle's "Unmoved Mover") manifests itself only through a special "derived meaner." Or, epistemologically speaking, the latter endows the former with its "original intentionality," even though the latter derives its ontological being from the former.

(When one says, following Dennett, that our own sightful and

insightful powers of representation originate from a totally blind and unrepresenting source, one seems to lay one's finger on a paradoxical phenomenon, a phenomenon which is perhaps describable as successful bootstrapping, and whose amazing or "miraculous" feature may easily escape from our attention when it is simply subsumed under familiar generic terms like "creation" or "emergence.")

This observation of mine, if correct, may help show that there is some deep insight in the so-called Anthropic Principle in cosmology,[1] which roughly says that it is no accident that the cosmic process in which we find ourselves has exhibited such a miraculous degree of "meanings" (say, unity and simplicity such as coincidence of ratios between basic constants of macro/micro-physics); it is, in a sense, because we've "selected" it, or it is the only (actual) cosmic path we inherit or survive among countless possible other paths that either died out or hadn't got a chance. The cosmic fact we survived or emerged or were "selected" or whatever you call it, though itself a purely contingent brute fact, made all the difference, as far as the conceptual necessity is concerned, when reflective yet limited beings like ourselves try to make sense of the immense phenomenon of that unbroken string of triumphs in evolution.

What I am trying to convey here is perhaps a necessarily vague idea: somehow there is no alternative way of thinking insofar as the contingent cosmic process that ultimately ended up with our own appearance is the only process that offers patterns that are bound to be recognizable, at least in principle, by late comers like us. We don't even have any means to verify that such a grand, all-inclusive process itself is contingent (as opposed to "necessary" in some ordinary sense). So when we understand Darwin's claim that nature has "selected" us (with or among everything else), the best sense we could make of such "selection" is an ex post facto sense and in the only available perspective of one unique species of the "selected" whose intentional stance alone endows it with its purported meaning. (Perhaps not so trivially, the reason we readily understand the meaning of "natural selection" is that we are already so familiar with our own personal selections of various kinds.) Therefore the term "ex ante natural selection" seems to make less literal sense than that of "ex post

reverse projection." We could have never existed, but once we came into existence, it's necessary for us (i.e., we've got no choice not) to see, when relevant epistemic conditions obtain, the miraculous (i.e., extremely improbable) trace of natural selection, and to recognize as real intentional patterns or ascriptions to other minds as well as our own. Thus the notion of "normativity" obtains its anchorage at this deepest level.[2]

Notes for Chapter 11

1. There are different expressions of this principle. E.g., Oxford English Dictionary defines it as "the principle that theories of the universe are constrained by the need to allow for man's existence in it as an observer." It quotes B. Carter's expression (1974), "what we can expect to observe must be restricted by the conditions necessary for our presence as observers."

2. As a postscript, it might help just to enumerate what I take to be a few conceptual issues related to my thesis of deep normativity:
   a) The space of causes vs. the space of reasons
   Conceptually, causes are not reasons; they belong to different logical spaces. (There is no issue of incoherence between any causal items as natural occurrences, while there can be inconsistency between reasons. Or, nobody can violate natural/causal laws whereas one can easily violate logical rules as principles of reason.)
   In reality, causes (or their relations, patterns) can turn into reasons, or be appealed to as grounds for reasoning. Conversely, in the domain of action (as depicted by Davidson's theory), the (primary) reason for an action is its cause. (A question: to which space do causal laws themselves belong?)
   b) The status of teleological terms (such as goals, targets, purposes, functions).
   Genetically, such notions ought to be prior to the emergence of intentional terms/stance. (They are perfectly applicable to many organisms.) Perhaps they can serve as intermediates, or a bridging phase between causes and reasons.
   • Epistemologically, they depend on the intentional stance for their bona fide existence as teleological items.
   • When one thing (say, a litmus paper) "detects" or functionally

discriminates other things (say, acid or alkali), can we say that it "recognizes" them without presupposing the intentional stance?

c) Are mental states or attitudes ontologically constituted by (rather than merely epistemologically dependent upon) the intentional/theoretical stance or stance-adopting activities?

At every evolutionary level below that of human intentions, we may say that whatever is conveniently describable by the intentional stance can be replaced, without any loss of content, by shifting to physical plus design stances; whereas at the intentional level, such a no-loss-replacement seems impossible. This implies a thesis of the primacy of normativity (stance-taking) in explaining intentionality.

**[**The full references are listed at the end of the book.**]**