

PERSONS IN PROSPECT*

III. Love and Non-Existence

The birth of a child is an event about which we can have “before” and “after” value judgments that appear inconsistent. Consider, for example, a 14-year-old girl who decides to have a baby.¹ We tend to think that the birth of a child to a 14-year-old mother would be a very unfortunate event, and hence that she should not decide to have a child. But once the child has been born, we are loath to say that it should not have been born. We may love the child, if we are one of the mother’s relatives or friends, and even if we don’t love it, we respect it as a member of the human family, whose existence is not to be regretted. Indeed, we now think that the birth is something to celebrate — once a year, on the child’s birthday.

We may be tempted to explain our apparent change of mind as a product of better information. Before the birth, we didn’t know how things would turn out, and now that the baby has made its appearance, we know more. But the birth does not bring to light

*

The paper bears some similarity to Larry Temkin’s “Intransitivity and the Mere Addition Paradox”, *Philosophy and Public Affairs* 16 (1987): 138-87. Both seek to show that a combination of views about future persons is not as paradoxical as it seems. The difference between the papers is this. Temkin focuses on failures of transitivity among comparative judgments; I address a different problem, in which the value of a general state of affairs appears inconsistent with the values of all possible instances. I am unsure whether the metaethical solution that I propose for the latter problem is called for by the former.

The paper also overlaps in important respects with Caspar Hare’s “Voices From Another World: Must We Respect the Interests of People Who Do Not, and Will Never, Exist?”, *Ethics* 117 (2007): 498-523. In the last section of that paper, Hare discusses the difference between *de re* and *de dicto* concern for persons, which is more or less the same difference that I discuss here.

Finally, Jeff McMahan discusses many of the same issues in “Preventing the Existence of People With Disabilities”, in David Wasserman, Jerome Bickenbach, and Robert Wachbroit (eds.), *Quality of Life and Human Diference: Genetic Testing, Health Care, and Disability* (New York: Cambridge University Press, 2005), pp. 142-71. My approach to these issues is similar to McMahan’s in some respects and different in others. The closest similarity is to remarks that he makes about “attachments” on pp. 159 ff. The greatest difference is that McMahan analyzes cases of this kind as involving changes of evaluative judgment, whereas I analyze them as involving pairs of judgments that seem inconsistent only if understood in mistakenly realist terms.

¹ This case is discussed at length by Derek Parfit in *Reasons and Persons* (Oxford: Oxford University Press, 1984), Chapter 16.

any previously unknown information that is relevant to our judgments.² Or, at least, I mean to restrict my attention to cases in which it doesn't. There may be cases in which we disapproved of the girl's decision because we feared specific dire outcomes, such as a birth defect or a descent into juvenile delinquency; and then if such outcomes don't materialize, we may change our minds. I am not speaking of such cases; I am speaking of cases in which we disapproved of the girl's decision for reasons that are not falsified by subsequent developments, and yet we are subsequently glad about the birth. The child is indeed raised under serious disadvantages of the very sort that we anticipated, but even the severely disadvantaged child is a member of the human family whose existence we cannot bring ourselves to regret.

In such a case, we even knew in advance that our perspective would change. Even as we deplored the girl's decision, we knew that the child would be a person whose existence we would celebrate rather than regret. We may even have cited this fact to ourselves as an argument for changing our minds beforehand: "Don't condemn her for deciding to have a child," we might have said: "The child will be loved, and what could be wrong about bringing a beloved child into existence?" Or we might have been told, by someone whose mother had also been under-age at his birth, "If you condemn her decision, then you're implying that I shouldn't have been born!"³ Such arguments strike me as defective. But what if anything is wrong with them?

This conundrum is one of several that Derek Parfit considers in Part IV of *Reasons and Persons*, the part devoted to "Future Generations". I want to suggest a solution that Parfit doesn't consider. Parfit's entire discussion is predicated on the assumption that our value judgments must be consistent as descriptions of the things they evaluate: they must be satisfiable by some distribution of value across the possibilities. I think that the present

² McMahan makes the same point on p. 155.

³ Parfit considers this case, p. 364.

case gives us reason to reject this assumption.

I do not think that we actually change our minds after the birth of a child, if a change of mind would entail giving up our antecedent judgment. I think that we retain our antecedent judgment while adding another judgment, directed at a slightly different object. Our antecedent judgment was that the girl shouldn't have a child; our later, additional judgment is that the birth of this particular child is to be celebrated.

At this point, one might think that our value judgments can be reconciled, after all. What we thought beforehand was that the 14-year-old girl should not have a baby. The target of our judgment was a quantified proposition, expressing the possibility that she has some baby or other. What we think afterwards, however, is that the birth of *this* baby was a fortunate event. And of course we can consistently think that the state of her having *a* baby would be unfortunate in general but not in the case of her having this particular one, since the general rule affirmed by our first judgment may allow for an exception that is cited by the second.

Here again, however, the attempted reconciliation appears to be blocked by the fact, which was known to us in advance, that any baby she had would be loved, and that its birth would therefore be viewed as a blessed event, and rightly so. How can we judge that the girl's having a baby would in general be unfortunate, while also judging that any individual baby would be an exception to the rule? If any particular baby she might have would be such as should be celebrated, how can her having one of them be deplored?⁴

The way to resolve this conundrum — or, rather, the way to dissolve it — is to question the assumption that our evaluative attitudes and judgments must reflect some distribution

⁴ This combination of attitudes might make sense if, for example, we deplored the injustice of her having one of the possible babies rather than the others. But the reasons for which we actually deplored her having a baby would not account for the attitudes in question.

of value across possibilities. And the way to question that assumption is to adopt a conception of value as a projection of rational attitudes rather than an independent reality to which they respond.

How could it be rational to have such different attitudes toward the birth of a baby? The answer lies in the difference that I have already noted in the objects of these attitudes.

Our later judgment is *of the child* in an especially immediate way, because it involves thinking of it under a demonstrative mode of presentation, as “this child”.⁵ We can make such judgments because we are acquainted with the child and can therefore have modes of presentation that pick it out demonstratively. When his mother was still deciding whether to have a baby, no one was acquainted with the actual baby that she would have, and so no one could not make judgments *of it* in this sense.

Why does it matter whether we can make judgments about someone considered demonstratively? The reason is that such judgments can be guided by emotions that similarly pick out their objects under modes of presentation derived from acquaintance. One such emotion is love. In the context of its mother’s love, the child is presented to her mind as it has become familiar to her, available for mental demonstration as “this baby”, “him”, or “her”. She does not love it under less immediate modes of presentation like “such and such a child” or even “Fred” or “Sue”. The latter modes of presentation would have been available to her even if she had merely heard the child described or referred to by name, in which case she would be in no position to feel love *for it* in the appropriate sense. Unlike these modes of presentation, acquaintance-based thought is way of being

⁵ Here I mean, not just that our name or concept doesn’t pick him out by an implicit description, but that its reference wasn’t fixed by a description, either.

mentally in touch or *en rapport* with an object; and the rapport it entails is prerequisite to the emotion of love.⁶

Our mental rapport with something can determine which attitudes toward it are rational to have. Loving a baby before being acquainted with it not actually possible, in my view, hence not rational, either; loving a baby after being acquainted with it is the easiest thing in the world — rational, too. But the general state of affairs consisting in the girl's having a baby is accessible only to attitudes like desire and approval, or aversion and disapproval, rather than an attitude such as love.

Our value judgments about this case seem to conflict because they correspond to the different responses that are rational to have toward it as we grasp it under different modes of presentation. We should feel free to experience these responses and hold the corresponding judgments, because value is the shadow of such attitudes, not an independent standard of their correctness. If the attitudes make sense, then the fact that they cast conflicting shadows cannot undermine their authority. And they make sense, despite the conflict between their shadows, because their intentional objects are different in ways that rationally affect the emotions that inform our judgments.⁷

⁶ Thus, an expectant mother who says that she already loves her future child cannot be speaking the truth, in philosophical strictness. She may be imagining how she will love the child, mentally simulating what it will be like to love the child, or having fantasies of loving it. But until she becomes acquainted with it, her love cannot be *for it*, and so the emotion she feels cannot be love.

When does a prospective mother become acquainted with her child? I would say that she becomes acquainted with it when she first perceives it. And when does she first perceive her child? I would say that she perceives the child's body, if not the child himself, at the point traditionally called quickening, when the fetus begins to make movements that she can feel. Thus, the tradition that interpreted quickening to be a morally relevant threshold was not just a superstition, in my view; it drew what may indeed be a morally relevant distinction)

⁷ This dissolution of the problem would be unnecessary if our emotions led us to judgments positing distinct and incomparable values. If we judged merely that the girl's initial decision was imprudent, whereas the baby is beautiful, then we could interpret our judgments as descriptions satisfiable in the one and only actual world, on the grounds that beauty has nothing to do with prudence. Pluralism about values could thus spare us from resorting to anti-realism.

My explanation of these value judgments raises questions about a term that figures prominently in Parfit's work and the literature that it has spawned. The term is "a life worth living". I believe that there is no coherent concept attached to this term.

Ordinarily, when we ask whether a life is worth living, we are asking whether it is worth continuing. Our question is whether the benefits of continuing to live will adequately repay the subject for the associated burdens. The problem with even this ordinary sense of the term is that it gestures toward a comparison where comparison seems impossible. For whether the benefits of continued life are worth the burdens must depend on the alternatives: any balance of benefits to burdens may in principle be worthwhile if the all of the alternatives would entail less of one and more of the other. But in the case of continued existence, there is no balance of benefits and burdens with which to compare, since the alternative is nonexistence, in which there would be no subject to whom benefits or burdens could accrue. How, then, do we tell whether life is worth continuing?

This problem is easily solved, however. When we ask whether a life is worth continuing, we are asking whether the subject has good reason to go on; and such a reason would consist in some possible life-event whose inclusion in his life would make it better as a whole, better than if he had not lived to see it. When someone wants to live

But I am imagining us as drawing — as I think we do draw — all-things-considered conclusions about whether a baby, or this baby, should have been brought into existence. And I am imagining that, whereas we still think that the girl shouldn't have had a baby, we think otherwise about this baby's having been had. Pluralism about value won't render these judgments compatible.

Parfit considers other ways of dealing with the conflict, but none strike me as satisfactory. For example, Parfit claims that, were he the child of a birth that was unfortunate when viewed prospectively, he would agree in retrospect that he shouldn't have been born. At this point, I think, we hear the grinding sound of bullets between teeth. I would rather think that it makes perfect sense to have different attitudes toward a child we know by acquaintance than we have toward a child considered abstractly, even if those attitudes cannot be modeled by a single distribution of value.

I also prefer this solution to the one favored by McMahan, according to which we change our minds about the girl's decision to have a baby. McMahan considers a solution like mine, when discussing the evaluative import of "attachments to particulars", but he ultimately drops the solution in favor of one based on a change of mind.

long enough to finish an important project or have some meaningful experience, he probably thinks that doing so would help to complete his life or bring it closer to perfection.⁸ And in that case, he is making a comparison for which the requisite alternatives are to hand — namely, his life extended to include the valued event, on the one hand, and his life cut short without including it, on the other. He can consider whether the one life would be a better life to have lived than the other. If the answer is yes, then he will say that the possibility of the event in question gives him “a reason to live”.

But a reason for him to live, in this sense, is not necessarily a reason for him to be glad that he was born. Having started an important project, he may judge that he will have lived a much better life if he finishes it than if he dies leaving it unfinished; and yet he may not think that his finishing the project will justify his very existence, since the value of finishing the project may be contingent on his having started it. If he had never existed to start the project, his not existing to finish it would have been neither here nor there.

Unfortunately, Parfit uses the term “a life worth living” in the latter sense, meaning a life that there is sufficient reason to have had from the beginning. And this sense of the term gestures toward a truly impossible comparison rather than the one indicated by the other use of the term. A person cannot compare the balance of benefits and burdens in his life to those of never having lived at all. The latter possibility would really have been neither here nor there, since he would not himself have been anywhere.

Parfit has a solution to this problem. His solution is to ask whether the person would regret having been born. According to Parfit, the subject’s preference (actual or ideal)

⁸ See Bernard Williams’s discussion of “categorical desires” in “The Makropoulos case: reflections on the tedium of immortality” in *Problems of the Self* (Cambridge: Cambridge University Press, 1973), 82 - 100.

between his existing and his never having existed determines whether his life is worth living.⁹

Here is the point at which Parfit runs afoul of the difference in judgment that can be wrought by acquaintance-based thought and its associated emotions. As I have argued, we can consistently judge that the birth of a child to a 14-year-old girl is regrettable but not the birth of this child — or that child, or any other demonstratively specifiable child. The problem for Parfit is that the baby itself will be able to judge likewise when it grows up. The future adult will be able to think that no baby should have been born into a life such as his, while still being glad that *he* was born. He can have both attitudes because the former is an attitude toward a general state of affairs while the latter is an attitude toward himself considered demonstratively as “this person” or “me”. Indeed, the inescapability of valuing himself from the latter perspective may lead him to regard his situation as ironically cruel, since it is of an undesirable sort whose instances are nevertheless irresistible to their occupants.

I think that the same conflicted feelings can arise in the parents of children who are born severely disabled. These parents are, so to speak, doomed to love a child such as is regrettable to have — regrettable, that is, when considered as such a child, not of course as this child in particular. In this respect, the parents are caught in a bind partly created by their love for their own child, a bind of a sort that makes the birth of such a child all the more tragic. Similarly, a child born into unfortunate circumstances is doomed to be attached to a particular existence such as is regrettable to have. As an adult, he may resent the fact that his inevitable self-attachment forces him to be thankful for having been

⁹ See p. 487: “[A person] might . . . decide that he was glad about or regretted what lay behind him. He might decide that, at some point in the past, if he had known what lay before him, he would or would not have wanted to live the rest of his life. He might thus conclude that these parts of his life were better or worse than nothing. If such claims can apply to parts of a life, they can apply, I believe, to whole lives.” In my view, Parfit here misinterprets the comparison that is made by someone who regrets having continued to live after some point in the past. According to Parfit, the person is judging his life since that point to have been “worse than nothing” — worse, that is, than nonexistence. I would say that the person is judging his life *with* its recent continuation to be worse than the life he would have had if he had died instead.

given a life of such an unfortunate kind.

The question is whether that adult will regret his own birth in the sense that determines whether his life is worth living, according to Parfit. He regrets the birth of a child into such a life but not his own birth. So is his life worth living or not?

My sense is that Parfit wants to give the benefit of the doubt to lives whose subjects are or would be thus conflicted. That is, he judges that a life is not worth living only if the person living it would, when thinking of himself demonstratively, regret that he was born. The result is that Parfit takes sides with the inevitable self-attachment that a person feels for himself considered as “me” — the very attachment that may force him to be thankful for an existence that he thinks undesirable for anyone to have.

My own inclination is to see this preference as rather cruel. I am inclined to say that we should not bring people into lives that they can be thought of as doomed to cherish or constrained to be thankful for. In any case, we cannot assume that there is a fact of the matter as to which criterion of regret we should apply when judging whether lives are worth living. Hence we still lack a determinate comparison that would give a clear meaning to the term “a life worth living”.

My explanation of our value judgments also bears on the problem that dominates Part IV of Parfit’s book, the so-called non-identity problem. In the case of the 14-year-old girl, the non-identity problem is supposed to be this: If she had postponed motherhood until she was older, she would have had a different child, and so her actual child cannot claim to have been harmed by being born to an underage mother, since *he* could not have been born when his mother was older. How, then, has she done any wrong?

Parfit fleetingly considers what I believe to be the correct solution to this problem. The solution is that a child has a right to be born into good enough circumstances, and being born to a 14-old-mother isn't good enough. What is "good enough" depends, in the first instance, on an absolute standard set by basic human needs; but it also depends, I think, on what the child's parents could at some point provide to a child.

Because the prospective mother can think of her future child only under a description such as "my first child", she can be guided only by *de dicto* considerations about it.¹⁰ The attitudes that would warrant distinguishing between the different children she would have at different ages — attitudes such as love or attachment — are not available to her. The difference between having *this* child when she is 14 and having *that* one when she is 21 cannot make any difference to her, precisely because she cannot think of them as "this" or "that" child.

The prospective mother is therefore confined to deliberating in light of what she can provide for "my first child", considered *de dicto*. And in her deliberations, she is obliged to think how best to provide for her child so considered, a thought that militates in favor of having her first child later in life, when she will be better able to care for it. The thought that the later-born child would be a different individual is available to her in the abstract, but it is not available in the form that could engage the evaluative attitudes that, after the birth, will lead her to celebrate having had *this* particular child, no matter which child she has. And her present obligations are limited to considerations that can rationally figure in her deliberations. Her obligation toward a future child is therefore what Caspar Hare calls a *de dicto* obligation, the obligation to provide a better start in life for her first child, whoever it turns out to be. If she provides her first child with a significantly worse start than she could, she will have violated an obligation to it. Her first child has a right to something better.

¹⁰ The arguments of the following paragraphs draw on Caspar Hare's notion of *de dicto* obligations. See Hare's "Voices from Another World".

Keep in mind that, as I argued in Part I (“The Gift of Life”), bringing a child into existence is, in Seana Shiffrin’s phrase, an “equivocal act”, because it throws the child into a predicament.¹¹ A human child is born without the capacities whose exercise will constitute its flourishing: it must develop those capacities, plus an identity with which to make sense of investing them in specific pursuits. The child also faces the distinctively human tasks of mastering its freedom and befriending its body, and all of these tasks will require years of work. Refusal to take them up, or failure at carrying them out, will entail an inability to cope with the exigencies of human life; and as I argued, being unable to cope is constitutive of suffering. When considering whether to throw a child into this predicament, in which it will be not yet equipped to flourish but all too well equipped to suffer, one must consider oneself obligated to do so only under conditions that maximize the assistance that one can provide — a level of assistance to which the child will have a right at birth.

This birthright has already been violated in the case of a child born to a 14-year-old mother. Parfit initially seems to think that the impossibility of fulfilling the birthright of this particular child somehow excuses its mother from any obligation to do so. He then realizes, however, that the mother will be guilty of having created a child whose birthright is impossible to fulfill, which may be just as bad as not fulfilling that right. In this respect, her decision to have a child is like the decision to give a promise whose fulfillment is already out of the question. The latter is also a decision to create an unfulfillable right, and it is clearly wrong.

Yet Parfit dismisses this solution to the non-identity problem on the grounds that the child, being glad that it was born, is likely to “waive” its birthright. Since the unfulfillable right she creates is bound to be waived, he thinks, the mother is off the hook. Here

¹¹ Seana Valentine Shiffrin, “Wrongful Life, Procreative Responsibility, and the Significance of Harm”, *Legal Theory* 5 (1999): 117-148.

Parfit's reasoning is confused in two ways, one of which involves the nature of acquaintance-based value judgment. (I'll discuss the other confusion in a footnote.)¹²

The fact that a child would be glad to have been born does not entail that he would be inclined to waive his birthright. His birthright belongs to him as *a* child, not as this particular child; but it is this particular child whose birth he is glad of. He can reasonably say to his mother, "I'm glad that *I* was born, but you were wrong to have a child whose rights you could not fulfill." Not only can he reasonably say this; he probably *will* say it, once he realizes that other children have been given, and sensibly regard themselves as entitled to, the best start in life that their parents could provide to a child. He will continue to assert his birthright, despite being glad that he was born.

My attempt to vindicate these seemingly inconsistent judgments depends on the claim that they are based on a rational pair of attitudes. Yet the attitudes themselves may seem irrational precisely because they support inconsistent value judgments. How can it be rational for a person to be glad, all things considered, about his mother's having done something that he regards, all things considered, as regrettable?

¹² Ordinarily, the prospect of waiving a right arises in the context of three possible outcomes. We can: (1) retain the right in order ensure either (a) that it is fulfilled or (b) that we will have legitimate grounds for complaint if it isn't; or we can (2) waive the right. Considering all three outcomes, we may prefer to retain the right, even though we would prefer to waive it if outcome (1)(a) were excluded. That is, we may think that retaining the right for the sake of possibly having it fulfilled would be sensible, but that retaining it merely for the sake of having grounds for complaint would be petty and foolish.

Given our preferences, the party against whom we hold the right can induce us to waive it if he can manage to take outcome (1)(a) off the table. But surely a waiver obtained by such means would not be normatively valid. He cannot gain release from fulfilling our right by confronting us with the fact that he isn't going to fulfill it, so that our only alternative to waiving the right is to retain it for the petty purpose of lodging a complaint.

To be sure, the child of a 14-year-old mother cannot exactly claim that she has taken outcome (1)(a) off the table: it was never *on* the table for this particular child. And yet the child may still waive his birthright because his only alternative is to complain that it cannot be fulfilled. And such a waiver is granted less voluntarily, because it is granted in the presence of fewer relevant alternatives, than the waiver of a right that can still be fulfilled. Its validity is therefore questionable.

How, in fact, can these attitudes be assessed for rationality at all, if not in light of a distribution of values to which they are answerable? Surely, what's rational by way of gladness depends on what's worth being glad about, which is just whatever has a particular type of positive value; whereas what's rational to regret depends on what's regrettable, which is just whatever has a particular type of negative value. And as I have already conceded, there is no consistent distribution of values under which the general fact of bearing a child has the relevant negative value even though any one of its instances would have the relevant positive value. On what basis, then, can I claim that the corresponding attitudes are rational?

Let me outline a conception of value that supports this claim. I'll start with the relation between value and evaluative response.

There are people whom I like despite knowing that they aren't very likable, and then there are people whom I know to be likable even though I just don't like them. Similarly, there are some jokes that I laugh at while judging that they aren't funny, and other jokes in which I can see the humor without be at all disposed to laugh. But when I say that I *find* someone likable, or *find* something funny, I suggest that I am doing some third thing. I am not just liking or laughing; I am discovering — “finding” — some quality that merits a response. But I am not simply judging that the relevant quality is present; I am finding it with the relevant sensibility, precisely by responding. I am detecting likeability or humorousness with the appropriate detector, namely, liking or laughter.

To find someone likeable or admirable or enviable, to find something interesting or amusing or disgusting — these are what might be called *guided responses*, responses that are somehow sensitive to indications of their own appropriateness. Guided responses are not value judgments or evaluations, since they are still conative or affective rather than cognitive attitudes. But they resemble judgments in being regulated for appropriateness, and so they are more than mere responses. Simply liking a person is a brute response that

does not yet amount to valuing him as likeable; valuing him as likeable is a conative attitude that does not entail passing judgment on his likeability. It is rather a matter of liking him in a way that is sensitive to what makes him worth liking.

But if anything makes someone worth liking, then it qualifies as a reason for liking him. Finding someone likeable is therefore liking him for a reason. What makes for the guided response that amounts to valuing, in other words, is guidance by reasons. And conversely, the reasons for valuing something are those considerations whose regulatory influence would make the difference between merely responding to it and valuing it — between, say, liking someone and finding him likeable, or laughing at something and finding it funny.

If my next step were to say that reasons for liking someone consist precisely in his likeability — that reasons for valuing something, in general, consist in its value — then my analysis would be fairly pointless. No philosophical work would have been done, since value is the term most in need of analysis. My aim is to fill that need, by proposing the opposite order of constitution. Something's value, I want to say, consists in there being reasons for valuing it, which are considerations whose regulatory influence would turn a brute response to it into an instance of valuing. Humorousness is that whose detection amounts to finding something humorous rather than merely laughing at it; likeability is that whose detection amounts to finding someone likeable rather than merely liking him.

The question, then, is how responses are regulated when they are more than casual or haphazard. What is guided laughter or guided liking?

Experimental psychologists have shown that we actually do regulate our responses in accordance with an identifiable set of conditions. We tune our responses so that they make sense in light of our conception of ourselves and our circumstances.

In one experiment,¹³ investigators asked subjects to report, at intervals following exercise, whether they still felt its excitatory effects. By measuring the subjects' levels of physiological excitation at the same intervals, these investigators detected an initial phase during which the effects of exercise continued and were perceived as continuing; a second phase during which the effects of exercise continued but were not perceived as such; and a third phase during which these effects had disappeared both objectively and subjectively. During each of these phases, erotic materials were shown to one third of the subjects, who were asked to report their degree of sexual arousal. Subjects exposed to erotica during the first phase reported no greater arousal than those exposed during the third phase; but those exposed during the second phase reported greater arousal than the others. Thus, arousal that was not attributed to exercise appears to have been misattributed to the erotica.¹⁴ The converse effect has also been demonstrated: subjects are less likely to display an emotional response if they have been given a physiological explanation for its symptoms. For example, shy people placed in a socially awkward situation do not feel or act shy if they have been told that they are being exposed to an invisible stimulant that tends to cause jitters and sweaty palms.¹⁵

How does this attribution mechanism work? Attribution theorists generally explain it in terms a drive toward self-understanding — or, as they prefer to say, toward “cognitive consistency”. This cognitive drive gives us a strong incentive to react in ways that we can explain in light of the circumstances, and to behave in ways that we can explain in light of our reaction. Feeling stirred, we look to our circumstances to suggest an interpretation,

¹³ Cantor, J.R., Zillman, D., & Bryant, J., "Enhancement of experienced sexual arousal in response to erotic stimuli through misattribution of unrelated residual excitation," *Journal of Personality and Social Psychology*, 32 (1975): 69-75.

¹⁴ The classic experiment in this line of research was Schachter, S., & Singer, J.E., "Cognitive, social and physiological determinants of emotional state," *Psychological Review* 69 (1962): 379-99. Schachter and Singer showed that subjects who had unwittingly received a stimulant of some kind were more likely than others to display emotion, and which emotion they displayed could be manipulated by cues from the environment. Subjects who had received a stimulant became either giddy or angry, depending on which emotion was suggested to them by the behavior of a confederate of the experimenters, posing as a fellow subject.

¹⁵ Brodt, S.E., & Zimbardo, P.G., "Modifying shyness-related social behavior through symptom misattribution," *Journal of Personality and Social Psychology* 41 (1981): 437-49.

and we then behave accordingly. In doing so, we can shape an inchoate disturbance into a specific response, or transform one response into another.

Initially we may feel excitations that could be symptomatic, say, of either nervousness, fear, or awe. Which of these responses we interpret ourselves as having depends on which response would make sense to us under the circumstances; how we go on to behave depends on how it would make sense for us to behave, given the response we interpret ourselves as having; and we thereby give our initially ambiguous feelings the stamp of either nervousness, fear, or awe, depending on which would maximize the overall intelligibility of situation, self, and behavior.

Why do our excitations come to fulfill our interpretation of them? The main reason is that our actions feed back into their psychological sources both causally and conceptually. Fearful actions can turn our response into fear partly by shaping the response itself, in the way that smiling has been shown to affect our mood.¹⁶ Similarly, an inchoate agitation can crystallize into fear once we act on it as such by taking flight. And fearful actions can also help to constitute which response we are having, since part of what makes the difference between nervousness and fear is how it is manifested in behavior.

Having noted this mechanism for the regulation of our responses, we need look no further, I suggest, for the kind of regulation that turns our responses into valuations rather than brute responses, on the one hand, or value judgments, on the other. Just as behaving becomes acting when it is regulated by the agent's conception of what would make sense for him to do, so reacting becomes valuing when it is regulated by the subject's

¹⁶ See J.D. Laird, "The real role of facial responses in experience of emotion: A reply to Tourangeau and Ellsworth, and others", *Journal of Personality and Social Psychology* 47 (1984): 909-917; S.E. Duclos, J.D. Laird, E. Schneider, M. Sexter, L. Stern, & O. Van Lighten, "Emotion-specific effects of facial expressions and postures on emotional experience", *Journal of Personality and Social Psychology* 57 (1989): 100-108.

conception of what it would make sense for him to feel. The upshot of the analogy is that intelligibility is the criterion of appropriateness for valuation as well as action. The considerations whose influence turns reaction into valuation are reasons for valuing, and they turn out to be considerations of intelligibility. So the considerations that make something valuable, by providing reasons to value it, are considerations in light of which valuing it makes sense.

Here I may seem to have turned an obvious explanation on its head. The obvious explanation is that conditions make a response intelligible because they make it appropriate, whereas I have said that conditions make a response appropriate because they make it intelligible.¹⁷ I am well aware of reversing this explanatory order. I do so without apology, on the methodological grounds that it assigns to the explanandum that term which is more in need of explanation. We can explain our responses without invoking evaluative notions, whereas we have difficulty explaining the nature of values at all. If the former explanations can help to provide the latter, progress will have been made.

This methodology is especially helpful in accounting for the subtle shades of objectivity and subjectivity that our evaluative responses display. On the one hand, the conditions of appropriateness for a response appear to depend on the sensibility that is capable of it. What makes something appropriate to admire depends somehow on what an admiring sensibility is attuned to, which is what tends to elicit admiration from a sensibility equipped for that response. On the other hand, the conditions of appropriateness for a response cannot be read off the actual responses of the relevant sensibility. What's appropriate to admire isn't merely what admiring subjects actually do admire. But how can what's admirable depend on the responsiveness of an admiring sensibility without collapsing into whatever actually elicits the admiring response?

¹⁷ In "The Authority of Affect", Mark Johnston argues that the positive or negative affect involved in a desire can render its motivational force intelligible by presenting its object as "appealing" or "repellent". I am not speaking of intelligibility in this sense; I am speaking instead of the psychological-explanatory intelligibility of a response, in light of its role in a person's mental economy.

This problem comes in varying degrees. To begin with, some people just aren't likeable or admirable, and their lack of likeability or admirability seems to be independent of the subject's perspective. But then we allow for individual differences of taste, which entail that who is likeable or admirable for me needn't be so for you. Even these person-relative values seem to transcend the actual responses of the relevant persons, however, since my likes and dislikes can fail to detect who is really likeable from my perspective. Then again, you and I can criticize one another's sensibilities as needing cultivation or refinement, as if there were an objective criterion of good taste. And yet different values appear to differ in their susceptibility to such a criterion, since we allow more leeway for tastes in liking than in admiration.

How can the conditions of appropriateness for a response be objective in some cases and relative to individual sensibilities in others, while also allowing for rational criticism of those sensibilities, and to different degrees for different reactions? The answer, I suggest, is that the general standard of appropriateness for a response is its intelligibility, which is determined partly by the psychological nature of the reaction itself and partly by differences among individual sensibilities, which can themselves be compared and criticized on grounds of intelligibility.

Consider what makes it intelligible to admire someone. Admiration has a distinctive functional role: it disposes one to emulate the admired person, to defer to him, and to approve of his words and actions. In acquiring these dispositions, one may become either more or less intelligible to oneself, depending on one's other attitudes — beliefs with which the person's opinions may harmonize or clash; aspirations in relation to which he may exemplify success or failure; interest in topics on which he may be an expert or an ignoramus; likes and dislikes of other people whom he may resemble.

If someone has been a washout in lines of endeavor like one's own, denies what one most deeply believes, specializes in what strike one as trivialities, reminds one of a hated

foe, and resembles no one else whom one admires, then admiring him would make no sense, and in two respects. First, it's hard to explain why one would acquire a disposition to emulate and defer to someone of that kind; and second, acquiring that disposition would make it hard to figure out how it made sense to behave. Would it make sense to emulate the person's failure in the very pursuits at which one otherwise hopes to succeed? Would it make sense to defer to his judgments contradicting one's deepest beliefs? These questions would have no clear answers, if one really came to admire him. In short, one would make less sense to oneself admiring him than not admiring him. That's why he isn't admirable, whether or not one admires him in fact

As this example illustrates, the criterion of appropriateness for a response is holistically interdependent with those for other responses, as are the corresponding values. Whether it makes sense to admire someone who excels at a pursuit to which one has hitherto been indifferent may depend on whether it makes sense to begin taking an interest in that pursuit — which may of course depend on whether it makes sense in other respects to admire the person.

Similarly, one's state of excitation may need to be diagnosed as either fear or awe or nervousness, but it is unlikely to be all three at once. What's frightening may therefore depend on what's awesome or unnerving, and *vice versa*. That is, what it makes sense to interpret as, and thereby resolve into, awe may depend on what it makes sense to treat as fear or nervousness instead.

Sometimes, different responses may be simply incompatible. Anger, ennui, and disgust tend to dampen amusement, and so it can be difficult to understand why we are laughing at things that would ordinarily offend, bore, or sicken us. We say, "That's not funny," though sometimes we are laughing as we say it; and then we may add, "So why am I laughing?" This rhetorical question confirms that the unfunny is that which we don't understand laughing at. The reason why we don't understand laughing at something is not

that it is unfunny; rather, we don't understand laughing at it because it's boring or offensive or disgusting — or utterly unlike the other things that amuse us — and the resulting incongruousness of laughing at it is the reason why we think it isn't funny, despite our laughter.

Thus, what it makes sense to be amused by depends in part on what it makes sense to be disgusted, bored, or offended by. And each of these latter responses has its own functional profile, determining how it fits into our self-understanding, perhaps in conjunction with yet other responses. What's admirable or desirable may therefore bear indirectly on what's amusing, by way of what is or isn't boring.

These examples illustrate, further, the idiosyncratic nature of responses and the corresponding values. What makes sense for me to admire is not necessarily what makes sense for you to admire, in light of the functional-explanatory connections between admiration and other responses such as belief, desire, love, hate, fear, and awe, in which you and I may also differ. Each of us can thus have sensibilities in light of which things can be valuable for one of us without necessarily being valuable for the other, because valuing them makes sense for one but not for both.

Idiosyncrasy has its limits, however. There are many responses that all of us tend to share by virtue of our shared human nature. Such nearly universal responses include: an aversion to pain, separation, and frustration; an inclination toward pleasure, connection, and the fluid exercise of skill; the inborn and automatic fight-or-flight response; an interest in the human face and form; an initial dislike of snakes, spiders, blood, and the dark; and so on. Given the holism of what makes sense in our responses, these fixed points of human nature constrain most if not all of our values. Some things are desirable for any human being, because desiring them will make sense for anyone; other things simply cannot be desirable, because desiring them won't make sense for anyone.

The fixed points of human nature place different degrees of constraint on the intelligibility of different responses. Disgust is directly plugged in to the physiological reactions of gagging and retching; desire is regularly sparked by the appetites, but it can also flare up independently, in response to just about any object; there may be nothing that human nature determines us to admire, and yet admiration is deeply embedded in the network of other attitudes; whereas amusement floats free of the network, except for the few connections through which it is inhibited by disgust, anger, and boredom. What makes sense by way of each response is consequently more or less constrained, depending on its degree of natural connectivity.

I believe that the previously noted shades of objectivity and subjectivity can be explained by these considerations — idiosyncratic differences in how it makes sense to respond, commonalities based in our shared nature or way of life, the possibility of responding incongruously and of cultivating more intelligible responses. As the intelligibility of a response is more closely tied to our individual characters, the response is susceptible to more specific guidance from a personal standard of correctness; as the intelligibility of a response is more closely tied to our shared nature and practices, the response is susceptible to more specific guidance from an inter-personal standard; and as sensibilities can be more or less intelligible in themselves, the standard of intelligibility in relation to them can be better or worse standards.

This account of values can also explain various other features that are well known to philosophers. Practical reasoning as I conceive it favors cultivating appreciative responses to things that belong to general kinds — kinds that are recognizable, if not by explicit description, then at least by family resemblance. Whatever makes it intelligible for me to laugh at a particular joke — thereby making the joke amusing, at least for me — would make it intelligible for me to laugh at any relevantly similar joke, which would

therefore qualify as amusing for me, too.¹⁸ Insofar as I can generalize about what kinds of jokes amuse me, or what kinds of people I admire, I can better understand why I am laughing at a particular joke or emulating a particular person.

Practical reason thus encourages me to identify kinds of joke, recognizable by family resemblance if not by description, that constitute what is amusing for me. It thereby pushes me toward a position that appears to confirm the view that being amusing-for-me is a real, descriptive property of things. The reason why amusingness-for-me comes to seem like a real property, however, is that I have cultivated a sense of humor that is regularly responsive to jokes of recognizable kinds, so that I can understand being amused, when I am amused. The same goes for my senses of admiration, inspiration, disgust, and so on: they have been cultivated under rational pressure to be responsive to recognizable kinds of things, which constitute what is admirable, inspiring, or disgusting for me.

Thus, the notion that values are properties distributed consistently among things or states of affairs is actually the reflection of a pattern into which our evaluative responses tend to fall when regulated in accordance with reasons for responding, which are conditions in light of which a response would make sense. The ultimate criterion of appropriateness for an evaluative response is intelligibility, which can be characterized independently of any postulation of values and can therefore be constitutive of values instead.

¹⁸ On a particular occasion, of course, the relevant similarity may not be an intrinsic quality of the joke itself: what makes it intelligible for me to laugh on this occasion may be that I'm drunk or nervous, which would make it intelligible for me to laugh at just about anything. Yet I am also under rational pressure to identify kinds of jokes that regularly tend to amuse me by themselves, so that I can comprehend my responses to jokes more generally, without reference to the circumstances. And a joke that's amusing for me on this occasion because I'm drunk or nervous may not be intrinsically amusing for me — not “really” amusing, I might say — because it is not the kind of joke that generally makes it intelligible for me to laugh.

Although the most intelligible responses are usually those which are consistent across recognizable kinds of things and coherent with our other responses, departures from this pattern can be more intelligible in isolated cases. After all, intelligibility is an holistic matter of overall explanatory coherence, which sometimes requires tradeoffs between alternative marginal gains or losses. And because values are constituted by intelligible responses rather than *vice versa*, we should tolerate cases in which the most intelligible responses cannot be modeled by a consistent distribution of values: they are simply cases in which the normal pattern of intelligibility doesn't hold.

Consider again the parents of a severely disabled child. These parents may feel that if they truly love their child, as they unquestionably do, then they cannot lament the fact of having had a disabled child; and yet they cannot help lamenting what is unquestionably a lamentable fact. The resulting sense of emotional dissonance can wreak additional damage on the child and the family. In my view, however, there is no dissonance between the emotions themselves; the dissonance is between values that the emotions are mistakenly taken to reflect.

The parents experience their emotions as assessing the value of the child's existence, an interpretation under which the emotions cannot be simultaneously correct. But their emotions make perfect sense despite there being no consistent distribution of values that they would reflect, because emotions can only project value, and only when appearing to do so enhances their intelligibility.

The parents should therefore forget about evaluating their child's existence and feel the emotions that clearly make sense for them to feel. What's intelligible in their responses may cast an inconsistent set of shadows on the world, but they are, after all, only shadows.