

Soritic Thinking, Vagueness, and Weakness of Will¹

Stephen P. Schwartz

Ithaca College

Abstract: Soritic thinking based on reasoning that is involved in the sorites paradox plays a crucial role in some forms of weakness of will. Such soritic reasoning leads to failures of behavior, but cannot be shown to be irrational by standard means. Thus weakness of will appears to be rational, whereas strength of will is irrational when viewed soritically. The puzzle is how to undermine weakness of will and expose it as irrational. Even though such weakness of will is not moral, moral-type reasoning involving the principle of equality can be brought to bear. Weakness of will can also be seen to be analogous to free-rider problems and the prisoner's dilemma.

Key words: weakness of will; vagueness; sorites paradox; reasoning; free-rider problems; prisoner's dilemma

1. Introduction

Logical paradoxes are not supposed to be relevant to the psychology of everyday life. The liar's paradox in the hands of Gödel helped to undermine programs in the foundations of mathematics, but has had no discernible impact on our daily lives. Russell's paradox upset Frege's project and led to major developments in set theory but not much else. Logical paradoxes are supposed to be interesting to specialists but not victimize the public. I have never heard of anyone in their everyday lives getting confused about the set of all sets that do not contain themselves, or being really troubled by self-referential propositions such as 'This sentence is false'.

Not so with the sorites paradox—the ancient and famous paradox of the heap—as I shall argue. The sorites paradox unlike the other logical paradoxes actually leads to confusion and defeat, and plays a crucial role in some forms of weakness of will. The sorites paradox is not a problem just for logicians. People who have never heard of the sorites paradox, and even those who have, are baffled by it in their lives. We are prone to engage in what I call 'sorbitic thinking' or 'sorbitic reasoning' and thereby cause ourselves

¹ I would like to thank Eric Barnes, Richard Creel, Rodmon King, Steven Lee, Michael McKenna, and two anonymous reviewers for this journal for helpful comments on earlier versions of this paper.

much grief. Soritic thinking makes weakness of will more understandable and less paradoxical, and even to appear rational. It turns out that strength of will is more paradoxical than weakness of will.²

2. The Sorites Paradox and Soritic Thinking

Let us briefly recall the structure of the sorites paradox. If I have a heap of sand before me, and I gently remove one grain, I will still have a heap. The removal of a single grain cannot make the difference between heap and no heap. By simple and apparently valid forms of reasoning I arrive at the conclusion that one grain of sand constitutes a heap. There are myriad forms of the paradox and a voluminous literature on it (ably summarized and discussed in Williamson, 1994). All of the versions of the sorites paradox depend on the vagueness of the terms involved. Since e.g. ‘childhood’ is vague, there is no single second of one’s life that is the final second of one’s childhood, followed by the first second of adulthood or adolescence.³ (If there were a final second of childhood, then ‘childhood’ would not be vague but would be absolutely precise.) But if there is no specific final second of childhood, then each second of childhood is followed by another second of childhood (if it is not followed by another second of childhood, then it was the final second of one’s childhood—contrary to the hypothesis). Here is a logical problem. If each second is followed by another, that is an infinite series.

We could stipulate a precise meaning for ‘childhood’ and define it as terminating at a definite number of seconds after birth (which alas is itself vague so that also would have to be precisely defined). This precise term would not be our term ‘childhood’ and

² Richard Tuck (1979) argues that free-rider problems are all forms of the sorites paradox, but he does not connect the sorites paradox with weakness of will or other personal problems. I discuss free-rider problems below. Chrisoula Andreou has also connected sorites-type reasoning with behavior but not specifically with weakness of will. Her interests tend more toward problems with addiction, free-rider problems and the environment. See Andreou, 2005 and 2006.

³ We are assuming that one is living a normal extended human life. If one dies in childhood, then there is a final second of one’s childhood.

would not be nearly as useful and would lead to problems of its own. For example, two different individuals who were otherwise indiscernible in their level of maturity would differ in that one would be a child but the other would not because she is some very short period of time older. In technical jargon explained below ‘childhood’ tolerates aging of one second. For the most part vagueness is ineliminable and essential, and thus the sorites paradox persists still unsolved and probably unsolvable.⁴

Consider two cases of action reasoning with clear similarities to the sorites paradox, and that involve what would standardly be called weakness of will. Here we see Ralph heading to bed after a lovely evening at the pub or philosophy conference. As he ambles upstairs toward his bedroom he passes the sink and remembers that he is supposed to floss his teeth every night before he goes to bed. This makes a lot of sense. Ralph wants very much to keep his teeth and avoid dental problems and he believes that flossing is a necessary and perhaps also in the circumstances a sufficient condition for achieving his goal of dental health. Since we accept some version of the Humean model of action motivation, we expect Ralph to floss his teeth.⁵ But wait, now is when the

⁴ There are those, such as Williamson, who claim that vagueness is only epistemic. There is a precise cut off, precise to the nano-second, of each individual’s childhood. For various reasons, the cut offs are unknowable and thus the appearance of vagueness. Aside from the implausibility of the claim that there is a precise final nano-second of one’s childhood, it would have no material effect on the points to come. Unless the agents in question have reason to believe that vagueness is only epistemic, they will be operating under the natural assumption that vagueness is real and thus their soritic thinking would be unaffected. I imagine that very few of us believe in the precise and unknowable cutoffs that Williamson favors. If some of us do, then my remarks are relevant to the rest of us.

For an eloquent defense of the claim that the sorites paradox will not admit of a genuine solution see Schiffer (2006). In any case, nothing rests on the claim that the sorites paradox is probably unsolvable. Many have thought themselves to have solved the paradox, but convinced few others. In any case, a formal solution would not affect the soritic thinking of non-logicians. Soritic thinking would still be appealing on an intuitive level. See the quote from Hampton below.

⁵ The Humean model views actions as motivated by desires and beliefs. Roughly, if the agent desires a goal and believes that an action that he can perform is necessary for achieving that goal, he will perform the action. Naturally there are many subtleties and difficulties surrounding the Humean model. We are assuming for the discussion here some plausible version of it without specifying it in more detail. The ensuing discussion could be adapted to non-Humean models of motivation as long as agents are viewed as

soritic thinking can come in. Ralph reasons to himself: ‘If I don’t floss tonight I won’t lose my teeth or damage my gums as a result of just this one non-flossing. Flossing or not flossing tonight can make no discernible difference to my dental health’. So Ralph neglects to floss and proceeds happily off to bed. This is a case of the sorites paradox in action, as it were. Just as the removal of one single grain of sand cannot delete a heap of sand, one single non-flossing cannot affect one’s dental health.

Ralph’s soritic reasoning seems to be sound, just as the sorites arguments are classically valid and have apparently true premises.⁶ But why not floss anyway? From whence does Ralph’s soritic reasoning get its grip on him? Unlike the leaving or removal of a single grain of sand from a heap, which nobody would care about, a single flossing has practical disvalue for Ralph. Recall that we are assuming Humean motivation. Ralph, just like the rest of us, is operating with a self-regarding calculus whereby given his beliefs and desires he is attempting to maximize his satisfaction of those desires. Herein lies the conundrum. According to Ralph’s calculus not flossing is the most reasonable course of action. Flossing is time consuming and somewhat unpleasant and the benefit to his dental health from this flossing tonight is vanishingly small if there is

attempting to maximize, increase, or sustain satisfaction, benefits, happiness, or some other good for themselves.

⁶ Just for the record a sorites paradox can be set out as a series of inter-locking *modus ponens* which could look something like the following:

A collection of n grains of sand is a heap.

If a collection of n grains of sand is heap, then a collection of $n-1$ grains of sand is a heap.

Thus a collection of $n-1$ grains of sand is a heap.

If a collection of $n-1$ grains of sand is a heap, then a collection of $n-2$ grains of sand is a heap.

Thus a collection of $n-2$ grains of sand is a heap.

And so on until we reach $n=1$. This argument is classically valid and valid in many non-classical systems such as intuitionist logic. It also appears to be sound. Notice that if any premise is false (assuming say that we are starting with $n=10,000,000$), then exactly one of the 9,999,999 premises will be false which implies that ‘heap’ is absolutely precise in the sense that for some i , a collection of i grains of sand makes a heap, but a collection of $i-1$ grains does not. Of course, we know that the argument is not sound since it has a false conclusion. Thus the paradox.

any at all. On the other hand the bed and soft pillows are so inviting and pleasurable late of an evening. Given the choice between losing his teeth or going to bed right now and skipping his flossing, Ralph's self-regarding calculus would indicate not to go to bed. That, however, is not the choice he is now facing. The choice is between going to bed right now and doing something that is no fun and would have no measurable effect on his dental health. Unfortunately for Ralph, if his reasoning is sound tonight, as it appears to be, then it will be sound tomorrow night, and the night after, and so on. And just as the paradox says the heap cannot disappear as we remove individual grains of sand, and yet in reality it will, so Ralph's dental health will suffer if he persists in his soritic reasoning despite its apparent soundness. This logical paradox is a real practical problem for Ralph. He is being weak willed in the sense that he knows he is not sticking to behavior, intentions, and resolutions that would have long-range benefit for him and that would help to bring about results that he strongly desires.

Consider another example of weakness of will involving soritic thinking. Brenda is a smoker. She believes that her habit of smoking will eventually lead to serious and perhaps fatal lung disease and she very much desires to avoid serious and fatal diseases. Now we see her being tempted to smoke. How can she resist the urge to smoke? Will reasoning help her? Let's see. She reasons as follows: 'I've smoked regularly for some time now. Either I already have lung cancer or I don't. If I have it, then smoking this cigarette won't make any difference to my getting or not getting lung cancer. If I don't have lung cancer, then smoking this cigarette won't make me get it. Surely one cigarette does not give one lung cancer. Smoking this one cigarette here and now cannot make any measurable difference to my health'. The smoker can reason similarly to Ralph. 'Smoking this particular cigarette now cannot have any measurable effect on my health, whereas not smoking it will have a very noticeable and unpleasant effect on my present state of contentment'. So by a purely rational assessment of self-interested costs and benefits, apparently smoking this cigarette now is the rational thing to do even though Brenda believes that smoking increases her risk of lung cancer and she very much does not want to get lung cancer. She does not believe that smoking this one cigarette will have any measurable effect on her getting or not getting lung cancer. So she smokes it. If Brenda believed that smoking this very cigarette would give her lung cancer, she

would flee it like the plague. The problem for Brenda is that smoking this very cigarette right now is not in conflict with her primary goals and desires and will satisfy an immediate and urgent desire. If Brenda's reasoning is sound now, it will also be sound ten minutes from now when she wants another cigarette, and so on.

My claim is that soritic thinking as we have imagined in the examples of Ralph and Brenda is real and common. Obviously there are countless opportunities for people to reason soritically: 'Eating this one piece of cake won't make me fat', 'Skipping my workout today won't have any discernible negative effect on my physical condition', 'Squandering this \$100 won't have any noticeable effect on my financial condition', 'Not saving any money this week won't affect my retirement plans (especially given the uncertainties of the stock market)', etc. Soritic thinking can get a foothold whenever achieving an overall or primary goal involves a vast number of tiny steps over an extended period of time. The individual micro-actions involved in achieving the goal are like the grains of sand in a heap, and just as no particular grain of sand, or even a small bunch (also vague) of them, is essential to there being a heap, no particular micro-action is essential to achieving the goal. Any individual micro-action is dispensable, and so is the next one, etc.

What is the formal justification for calling the reasoning of Ralph and Brenda "soritic"? The sorites paradox is generated from the vagueness of the predicates or terms involved. How does vagueness operate in the reasoning of Ralph and Brenda? The key paradox-generating feature of vague terms is their 'tolerance'. The idea is that vague predicates will tolerate small changes in the objects to which they are correctly applied. These small changes when iterated "under the screen" of relevance to the predicate lead to paradox. The claim of central writers on vagueness such as Michael Dummett (Dummett, 1975), Crispin Wright (Wright, 1975, 1987), and Peter Unger (Unger, 1979a, 1979b)⁷ is that all ordinary vague predicates are tolerant, and furthermore that this tolerance is ineliminable without fundamentally changing the nature of language. Just for the record here is Crispin Wright's formulation of tolerance:

⁷ For an extended discussion of tolerance principles and the sorites paradox see Schwartz, 1989. For a contemporary treatment see Eklund, 2005.

A predicate is *prima facie* susceptible to [the sorites paradox] just in case it is *tolerant*; that is, sufficiently small variations in some associated parameter are apparently insufficient to affect the justice with which it can be applied to something, whereas sufficiently large variations are always so sufficient. (Wright, 1987)

Now we can highlight why the reasoning of Ralph and Brenda is ‘soritic’. Although they would not put it this way, they view their situation as tolerating some divergence from their long-range goals without having any negative effect. Ralph reasons that his dental health will tolerate not flossing tonight. Brenda reasons that her respiratory health will tolerate smoking this cigarette. And yet, here is the paradox, neither believes that “sufficiently large” divergences would be tolerated without severe negative effect. All soritic thinking in action takes this form: ‘My long range goals will tolerate this small divergence, but not of course large divergence’. ‘My long range goal of not being overweight will tolerate eating this extra large yummy piece of cake, but not vast numbers of such pieces of cake’ and so on. Thus the soritic thinking of Ralph and Brenda, and the rest of us when we are similarly weak, follows precisely the pattern of the sorites paradox as described by Wright.

Soritic thinking is so appealing because in these sorts of cases it is not obviously unsound, and indeed the error of reasoning in sorites arguments has not been satisfactorily dissected. We know there is an error because eventually the heap does disappear, the child does grow up, the smoker develops lung disease (or is likely to), the fortune is squandered, the eater gains weight, and so on. The difficulty is going from the apparently sound individual arguments to the clearly unsound aggregation of the arguments. The problem is how to use reasoning to go from the ‘sufficiently small variations’ to the ‘sufficiently large variations’. There is no way in classical logic to add sound arguments together and get an unsound one. If ten, twenty, or twenty million individual arguments are each sound, then conjoining them is classically sound. The challenge, thus, is to find some effective way to defeat soritic thinking. This is a real practical task for logicians. Mere exhortations will not work. Our soritic reasoners already know that they should floss their teeth, quit smoking, get regular exercise, and save for retirement. The question is how does this knowledge determine that Ralph must

floss tonight and that Brenda must not smoke this particular cigarette, and so on. How can the general knowledge of what one should do in the long run get a grip on this particular action now being considered, when the current (micro) action is neither necessary nor sufficient, nor even discernibly contributory toward gaining the long term goal?

3. Soritic Thinking, the Prisoner's Dilemma, and Morality

Since the soritic reasoning is classically valid, the application of formal logic to solve our puzzle is going to be difficult at best. Even if contrary to our expectations a solution to the sorites paradox is found, it is unlikely to have much effect on ordinary action motivation. Jean Hampton emphasizes this point in her famous article on free-rider problems.

Such vagueness... encourages people to reason in a way that has been associated with the 'Sorites' paradox.... But sorites-like reasoning is supposed to be a mistake,... There are, thus far, no uncontroversial proofs showing how it is fallacious, although even were one to be given, it still seems to be the kind of (fallacious) reasoning that people would find tempting,... (Hampton, 1987, 257)

Hampton is discussing, in this context, reasoning involving free-rider problems and individual contributions to collective goods, not weakness of will. We will briefly pursue this connection below.

Since there is little hope for a formal solution to soritic thinking perhaps, then, there is some factual error in the soritic reasoners' thinking. Maybe one of their premises can be shown to be false after all. Unfortunately their premises seem solid and almost irresistible, especially if couched in terms of tolerance. Ralph thinks 'My goal for my dental health will tolerate not flossing tonight' and indeed it will. What he thinks is correct. Likewise Brenda's respiratory health will tolerate smoking this one cigarette as she surmises. Nevertheless we should seek for some way to undermine these claims.

There is some statistic that used to be mentioned that each cigarette one smokes shortens one's life by six minutes, or something like that. Likewise maybe each flossing means keeping one's teeth six minutes longer. Given Brenda's self-regarding calculus living six minutes might well outweigh the pleasure of one cigarette and Ralph might well be willing to spend five minutes flossing to keep his teeth six minutes longer. The problem with this line of attack is that there is very little actual support for such statistics (see e.g. Tsevat et al 1991). Ralph would not be able to cite evidence that his not flossing his teeth tonight can have any statistically measurable effect on the length of time he keeps his teeth. Anyway, at some future time Ralph can do things that will make up for the lost time. He can floss twice a day, use an electric tooth brush, give up candy, and so on. Likewise in the case of smoking there are many confounders that make calculations of life expectancy hopelessly murky. Diet, urban smog, electrical power grids, radon in the basement, medical tests such as x-rays, and who knows what else might contribute to getting or not getting lung cancer. So given Brenda's epistemic situation it would be most unlikely that she could defend to herself in factual terms a belief that smoking this cigarette would actually shorten her life in any measurable way, and certainly not in any way that would be vivid enough to overcome her present urgent desire to smoke. In any ordinary case, such a belief would not be firm enough and present enough to affect her action. Furthermore the claim that smoking this individual cigarette shortens one's life by six minutes (on average), even if supportable, assumes a lifetime of heavy smoking. Brenda might well hope and perhaps believe that soon she will quit at a point in her life span where the damage of smoking will be erased over time. (See more on this below.) In any case we can assume along with most writers on the subject of free-rider problems that the participants do believe that their contribution or lack of it is negligible. If Ralph believes that his not flossing tonight could have a significant and deleterious effect on his dental health, then he is not engaged in soritic reasoning. Likewise if Brenda believes that this cigarette or any cigarette shortens her life by six minutes on average, and she values this six minutes, and so on, then she is not engaged in soritic thinking and her weakness of will cannot be explained by soritic thinking. Such restrictions may diminish the extension of soritic thinking, but not by much.

Ralph's and Brenda's problems with soritic thinking have an analogue on the national and global scale that harks back to Hampton's discussion of free-riding and collective goods. For example, the polluter might (and in fact sometimes does) reason as follows: 'Dumping this gallon of sludge in the river will have no detectible effect on the river, much less on the nation's or world's environmental quality, whereas disposing of it "properly" will have significant costs to me'. Problems like pollution, global warming, waste of energy resources, and so on are difficult to solve partly because they are fertile ground for soritic thinking (Hampton 1987, p. 257). No specific piece of e.g. non-polluting behavior has a detectible effect on the environment, nor any detectible positive benefit to the individual, and can often have significant costs. (See Sinnott-Armstrong 2005 for a detailed discussion of issues arising in this context.) If people did not have a personal economic motivation to turn down their thermostats and buy energy efficient cars, they would have to look hard for practical reasons to do it. The practical effects on the environment and energy supply of my turning down my thermostat would offer vanishingly little motivation even if I cared a lot about those things. Likewise there is something sound about the person's reasoning who declines to vote 'because my single vote will not make any difference to the outcome'. In fact it won't. Typically a single vote is like a single grain of sand in a heap. Similarly to Ralph considering whether or not to floss tonight, the voter's logic seems correct in isolation, but leads to disaster if aggregated over all or even many voters.

The person considering dumping the sludge in the river is in a different situation than Ralph, however, in that there is an ethical dimension to polluting. We can invoke a Kantian principle of universalizability when reasoning with the polluter, or rather the polluter should invoke such a principle to herself. Such a principle would effectively rule out the polluting behavior. We could also appeal to ethical universalizability in arguing against the negligent voter.⁸ On the other hand, there is no ethical dimension to Ralph's

⁸ We are assuming that the agents feel the pull of morality—that they are not simply amoral. Even within a Humean framework we can assume that agents, for the most part, want to be reasonable and deserving of praise. Sinnott-Armstrong (2005) raises difficult issues about the application of moral reasoning in such cases, but I am assuming these can be overcome in ways that he suggests.

behavior and reasoning. There is no local, national, or global environmental or political issue at stake. His reason for flossing, if he does, is and ought to be purely self-regarding.

4. Soritic Thinking and Weakness of Will

One of the advantages of recognizing soritic thinking is that it helps us to understand typical forms of weakness of will. Brenda is being weak willed when she smokes that cigarette and Ralph is being weak willed when he fails to floss. Of course, these single isolated lapses would just barely rise to the level of weakness of will, but I am assuming that Brenda and Ralph are tempted by soritic thinking to lapse over and over again and often succumb. Furthermore soritic thinking need not be the only factor contributing to Brenda's and Ralph's lapses. Perhaps soritic thinking by itself would not often lead to a complete behavioral collapse vis-à-vis long range goals, boritic thinking can work synergistically as it were with other well known factors that motivate deviations from behavior that we know would benefit us. These other factors in the absence of soritic issues might not be sufficient to motivate. Soritic considerations that impart an aura of reasonableness can give that extra "tang" to desires that would otherwise dissipate.

As standardly described weakness of will is very paradoxical. The agent desires the goal and believes it to be the best for him, he also believes that some action that he can perform is necessary or sufficient for achieving the goal, but he voluntarily fails to perform it. Initially the Humean model seems not to be able to accomodate such weakness of will—it appears to be impossible. Weakness of will has posed a paradox for the Humean model. We now see that this analysis is too crude, however, and does not apply to Brenda and Ralph even though they are being weak willed. Certainly Brenda desires to avoid lung disease more than she desires to smoke this cigarette, and Ralph

Although the example of the voter has some features in common with a prisoner's dilemma, it is not fully analogous, since the voter risks nothing by not voting. I am assuming, to repeat, that the single vote will not have any effect on the outcome, whether or not it is cast.

desires to keep his teeth and avoid gum disease more than he desires to go directly to bed teeth unflossed, but the major goals of avoiding lung disease and keeping one's teeth do not directly and easily transfer to the micro-actions here and now because of the soritic nature of the cases. Indeed, since there is no obvious unsoundness in the agent's soritic reasoning, the weakness of will seems the less paradoxical choice within a Humean model. In the cases of Ralph and Brenda strength of will would need explaining—strength of will is paradoxical, not weakness. So we see that in cases of soritic thinking, weakness of will is not an inexplicable loss of rationality and succumbing to temptation. Rather it is a kind of rationality. It is almost as though weakness of will, at least in typical cases that involve soritic thinking, is not the result of a failure of reasoning but a result of too much reasoning. Thus we are still left with something of a paradox—namely that reasoning and rationality lead to weakness of will and ultimately to defeat.

Is this really so paradoxical after all? Are we not familiar with situations where reasoning and rationality lead to weakness and defeat? If the hero stopped to reason, he would flee the scene. Strength of will and heroics often seem a bit irrational. The hero risks his life to save his squadron. That can seem admirable and at the same time very irrational. It need not be irrational, however. Heroics and even the sacrifice of one's life can make sense on a Humean model. If the hero's desire to save his squadron is greater than his desire to live on without them, then given his purely self-regarding calculus, what he does is rational. The problem with the successful quitting smoker and the persistent flosser (i.e. those with strength of will) is that what they do is inconsistent with their self-regarding calculus. Ralph's expected gain from each individual flossing is less than his expected gain from not performing that flossing. So why would he floss? We may think that the hero's desires are somewhat irrational, but given his desires his behavior is rational. The flosser gains nothing and loses a lot by flossing tonight. At the very least he gains nothing detectible, and certainly nothing that can outweigh his quite detectible cost.

The idea that weakness of will is the result of too much reasoning is borne out by recent work on weakness of will. Richard Holton in his article 'Intention and weakness of will' (Holton, 1999) offers a penetrating analysis of the phenomenon of weakness of will. He argues in detail, with references to convincing examples and empirical research,

that weakness of will is failure to stick to our intentions or resolutions.⁹ According to Holton, actors show weakness of will when they revise an intention or resolution as a result of a reconsideration that ‘exhibits tendencies that it is not reasonable for the agent to have’ (Holton 1999, p. 248). (Sometimes, of course, it is reasonable to revise a resolution, and then there is no weakness of will.) In the sort of example invoked by Holton the actor has made a resolution to floss every night, to quit smoking, or to exercise every day. In the event, they revisit their resolution with considerations that undermine it. Although Holton does not mention the sorites paradox or soritic thinking, typically these reconsiderations that exhibit unreasonable tendencies involve what I have been calling ‘soritic thinking’. Given the actor’s self-regarding calculus these considerations need not be irrational and acting on them need not be unreasonable in the circumstances. As we have seen it is actually reasonable. Nevertheless the actors exhibit weakness of will because they are too ready to reconsider and revise their resolutions. Holton claims that all cases of weakness of will demonstrate this pattern.

When is it reasonable to revise an intention? Certainly when the revision is based on new and relevant information, or when the original intention was formed hastily or when one was in an impaired state—fear, inebriation, exhaustion, etc. Reconsidering and revising intentions and resolutions exhibits unreasonable tendencies of the agent when those intentions or resolutions ‘...were expressly made in order to get over one’s later reluctance to act’ (Holton 1999, p. 249). This seems to be the situation with Ralph and Brenda assuming that they made a resolution—Ralph to floss his teeth every night and Brenda to quit smoking. The idea is that once soritic thinking gets started it leads to revision of intentions, because in the event the reasoning is actually correct. Accordingly, the failure of reasoning of Ralph and Brenda and other soritic reasoners is not anything internal to their reasoning. The failure is in allowing soritic thinking to get started in the first place. Revisiting and rationally reconsidering reasonable and beneficial resolutions is unreasonable and weak, according to Holton.

Can we state categorically, then, that soritic thinking always exhibits unreasonable tendencies, i.e. unreasonable tendencies to revisit and revise intentions and

⁹ Holton calls ‘resolutions’ ‘policy intentions’. I find ‘resolution’ to be more in tune with ordinary speech and less technical sounding than ‘policy intention’. Holton uses the term ‘resolution’ in Holton, 2004.

resolutions? If so, this would be a striking and perhaps unanticipated instance where engaging in reasoning is unreasonable.¹⁰ Holton's very appealing idea is that once one has examined the data, one's situation and prospects, and any other relevant information, and fixed on a reasonable intention or resolution to avoid some behavior or undertake a series of behaviors, it is unreasonable for one to revisit that intention or resolution at each or even any instance. The appeal is based on the nature and purpose of intentions and resolutions. The role of intentions and resolutions is to be action guiding and organizing. Fulfilling this role requires that they have a certain stability. If intentions are irresolute and fluctuating, they fail in their purpose.

Despite its appeal, I am not convinced that Holton's approach completely solves the problem of explaining the unreasonableness and weakness of soritic thinking. For one thing he gives us no clear answer as to why reconsidering intentions and resolutions that '...were expressly made in order to get over one's later reluctance to act' exhibits unreasonable tendencies. The reason that the role of intentions and resolutions will be undermined by reconsidering them at the time of temptation is that they are subject to soritic thinking. They may be subject to other undermining techniques as well, such as judgment shift,¹¹ but clearly soritic thinking is going to be one, if not the, major threat to intention stability. One cannot help thinking that an intention or resolution should be able to stand up to reasonable and properly considered examination at any point. If it cannot, then there seems to be something wrong with the intention. So if revisiting intentions with soritic thinking is a threat to intentions and resolutions that exhibits unreasonable tendencies of the agent, then it must be that soritic thinking is unreasonable. But this is what is supposed to be explained. So instead of explaining the unreasonableness of soritic thinking, Holton's analysis assumes it. There are other less serious problems as well. For example, whether and to what extent intentions and

¹⁰ Circumstances in which engaging in reasoning is unreasonable are not that hard to imagine. For example, if there is no time to think—if one must act now or lose the goal—stopping to reason would be unreasonable. In circumstances where one is very unlikely to gain new information, or where acquiring new information is too costly, would be circumstances in which continuing to reason would be unreasonable.

¹¹ Judgment shift is the revising of judgments based on one's present desire to act or refrain from acting. See Holton 2004 for a discussion of judgment shift.

resolutions are playing a role in Ralph's and Brenda's weakness of will is unclear. Ralph may not have quite decided firmly to floss every single night for the rest of his life. His thinking may have been more diffuse. 'I know I should floss, the dentist told me to, I bought the floss. I will try to floss most nights'. Likewise with Brenda, soritic thinking may interfere not with her keeping an intention not to smoke, but with her making the intention in the first place. 'I'll quit tomorrow'. (Or next week, next month, next year, after I get settled in my new job, etc.) The upshot is that Holton's analysis, while offering some help, does not offer a full diagnosis and antidote for soritic thinking.

4. A Solution to the Problem of Soritic Thinking?

A more promising approach to undermining soritic thinking is based on the sort of response that we would give to the polluter or the negligent voter. Consider in more detail how we might respond to the negligent voter. The negligent voter says 'My vote won't make any difference to the outcome, and voting is time consuming and involves effort I would prefer to spend otherwise'. Surely this is correct. We will not get very far arguing that perhaps her vote will make a difference. Rather our response should be along the following lines: 'You care about the outcome and you admit that the outcome will have an effect on your well-being and satisfaction and, more importantly, you support democracy and government of the people. If every eligible voter acted on reasoning similar to yours then your desires and goals would be frustrated. So in order for your goals and desires to be satisfied, you are depending on and expecting others to make an effort that you are unwilling to make. You could not be content for all the other voters to reason as you do and act on that reasoning. You are making an exception of yourself, treating yourself differently from others when there is no relevant difference between you and those others. This is a violation of the principle of equality, which is a basic and fundamental principle of rationality. 'Treat like things like'. So your decision not to vote is fundamentally unreasonable'.

The difficulty with importing this excellent reasoning to the cases of Ralph and Brenda is that the appeal to the non-voter is moral. It rests on the basic moral principle of equality of treatment of relevantly similar individuals. As mentioned above, no such moral principle applies to the typical cases of weakness of will based on soritic thinking. Nevertheless the basic idea should be extendible to soritic thinking in that the principle of equality is not restricted to moral contexts although it has powerful applications there. Non-arbitrariness is a *prima facie* requirement of all reasoning. If x is P and there is no relevant difference between x and y , then y is also P . To assign one to P but not the other would be arbitrary—a violation of the principle of equality. This principle of equality applies everywhere and to everything, not just in moral contexts to sentient beings. Ralph and Brenda are violating this principle of equality when they are being weak willed even though they are not making any moral error. Ralph would not be willing to universalize his reasoning when he fails to floss tonight. He would not be willing to say to himself: ‘I will not floss tonight, and for the same reasons I will not floss every future night’ even though the same soritic reasoning that applies tonight would apply any night. Likewise Brenda would not be willing to universalize her smoking of this cigarette. The issue in these examples is not as with the moral perspective the unwillingness to universalize to all other agents, but the unwillingness to universalize to their own future agency. Thus Ralph and Brenda are making unreasonable exceptions of the micro-behavior they are now considering, or put differently they are making unreasonable exceptions of their present selves.

We can reason with Brenda in ways somewhat similar to the way that we reasoned with the negligent voter. ‘You are about to smoke this cigarette, but you will be in this exact same situation many many times in the future. You are not now willing to commit to smoking all those cigarettes then, because you know this will have an unacceptable effect on your health. So if you smoke this cigarette now you are treating the present instance differently than you would treat those future instances. In other words, you are making an exception of the present instance, even though there is no relevant difference between it and any future instance. Thus you are violating the principle of equality. Thus you are being irrational’.

This strategy is promising because the soritic thinker is not willing to embrace the entire sorites series, but only a single premise of the series. The paradoxical sorites series would typically be generated by iterating a premise of the form: ‘If eating n pieces of cake have not made me fat, then eating $n+1$ pieces of cake will not make me fat’. Or ‘If smoking n cigarettes have not given me lung disease, then smoking $n+1$ cigarettes won’t give me lung disease’. (See footnote 6.) Generally speaking these premises are true, or if not true, at least not demonstrably false. When iterated the premises validly entail that if one is not already fat, then eating any number of pieces of cake will not make one fat. Likewise smoking any number of cigarettes will not give one lung disease. Our soritic thinkers do not believe the conclusions of sorites arguments, nor should anyone else (even though we cannot convincingly dissect the error involved). Consequently the soritic thinkers are unable to embrace the universalization of their soritic behavior, which in turn means that acting on the single soritic premise they have in mind is unreasonable. If the soritic thinker actually believed the full iterated sorites argument, then their acting on the single premise would not be unreasonable given their beliefs and desires. In that case they would believe on the basis of sorites arguments that e.g. smoking no matter how many cigarettes would have no negative effect on the health of their lungs, or that no matter how many pieces of cake they ate they would not gain measurable weight. Such a belief is itself unreasonable—but given their beliefs and desires they are acting reasonably. The point at hand is to show that normal soritic thinkers, such as Ralph and Brenda and the rest of, are being unreasonable given their and our beliefs and desires.¹² That is: they and we are being unreasonable from the Humean motivational perspective. And this can be done along the lines I have suggested. The point can be put succinctly: The sorites paradox relies on the principle of equality to progress from stage to stage. Since she is a child at n seconds of life, and there is no relevant (or even detectible) difference between her maturity at n seconds and $n+1$ seconds, by the principle of equality, she is a child at $n+1$ seconds. Soritic thinkers, such as Ralph or Brenda, are

¹² Some philosophers dismayed by the lack of any prospect of solution to the sorites paradox actually embrace the conclusions and thus deny that there are heaps, people, or any other ordinary objects. See e.g. Unger, 1979a, 1979b. I do not believe that the rest of us can or should embrace this “nihilism”, at least not in the sorts of contexts relevant here.

violating the principle of equality, because they are unwilling to make just that progression from stage to stage—even though there is no relevant difference between stages.

We can challenge the soritic thinker with a dilemma. Either make the progression (i.e. universalize the reasoning and behavior) or reject the present stage (i.e. don't claim that it is alright to smoke this cigarette) on pain of violating the principle of equality. You are not willing to make the progression. Thus you must, on pain of unreasonableness, reject the present stage.

5. Prospects for Success

Will this strategy work? I believe that it will in the sense that the soritic thinker can be convinced that she is being unreasonable. Just as the argument that she is making an unwarranted exception of herself is a powerful and compelling argument against the negligent voter, the full realization that one is unwilling to universalize one's behavior is a powerful argument against its reasonableness. Nevertheless in order to achieve this result many difficulties must still be overcome.

Since the strategy relies on something like Kantian moral reasoning, it shares many of the difficulties of Kantian ethics. For one thing, our anti-soritic strategy would seem to eliminate too many forms of behavior, or at least condemn them as unreasonable. This allows opportunities for excuses and judgment shifts. Often people do things on a whim or just for fun that are perfectly harmless, give pleasure, and are not the least unreasonable. How does our strategy condemn soritic behavior as unreasonable but not harmless indulgences, whims, and purposeless fun? Consider someone who is a moderate drinker who rarely drinks before five p.m. but decides every now and then to have a beer in the early afternoon just for pleasure. This may not be at all unreasonable, but he is not willing to universalize his behavior, and thus his decision would seem to violate the principle of equality. Nevertheless we have to ask what his reasoning is. Most likely his decision to have a beer on this afternoon is based on contextual

differences specific to this day. ‘I don’t have to work later today, I won’t be driving anywhere, I haven’t had a mid-day beer in several weeks, etc’. Also our moderate drinker may not have similar desires on most other days. I am assuming that these factors make for a relevant difference between today and other days. Decisions for behavior that is just for fun or on a whim should be based on relevant differences and contextual factors that single out the setting as particularly conducive to the behavior.

No doubt the soritic thinker will avail himself of similar considerations. ‘I really need this cigarette to get through this phone conversation with my mother’, or ‘I’m going to skip this flossing because I’m so tired tonight’ and so on. Whether or not the soritic thinker can be convinced that this is unreasonable will depend on context. Cutting off these maneuvers will depend on the details of the case. If we are assuming a long series of failures of will and similar excuses, then the soritic thinker would not be willing to universalize the reasoning. As a single, isolated instance the behavior would not be unreasonable and would hardly be called weakness of will as we discussed above. In the context of repeated failures it would be. Thus the full context of behavior needs to be assessed. The simple fact is that the smoker knows that continuing to smoke is a habit that has no future. The non-flosser knows that if he wants to maintain dental health there is no future in continuing not to floss. In the context of these beliefs, the smoker’s soritic thinking that leads to lighting up this cigarette is unreasonable; likewise the thinking of the non-flosser is unreasonable. In smoking this cigarette now, or not flossing tonight, the agents are unreasonably disregarding their own future agency, similarly to the way the negligent voter is disregarding other voters.

We hope that soon the soritic thinkers, just as the negligent voter, will change their behavior. Our negligent voter would find discomfort in simply looking at herself in the mirror and exclaiming ‘Yes, I am being unreasonable, and I am making an unwarranted exception of myself, but I am still not going to vote’. Our smoker would squirm to think of herself as unreasonable at every puff. Thus even within a Humean perspective appeals to morality and reasonableness can get a grip against soritic thinking. Unfortunately for our will power, although the appeal to reason is not completely impotent, it is not always compelling of behavior change. The smoker rather than not lighting up might prefer to think of herself as in the grip of an irresistible addiction and

simply cease reasoning about it. Not flossing might easily be rationalized by judgment shifts about the actual value of flossing, hopes for medical breakthroughs that will obviate the need to floss, and so on.

Despite its difficulties we are unlikely to find a better solution to the problem of soritic thinking than moral-like appeals to universalizability. This can be most clearly seen if we consider again the analogy of the types of weakness of will we have been considering with free-rider problems. The key feature of the appeal to the principle of equality is shifting the focus from the action here and now to the general situation and one's entire future agency. Thus smoking this cigarette now or not flossing tonight can be seen as free-riding on one's future selves' collective action. As with free-riding, their behavior can be considered irrational from a quasi-moral perspective involving the principle of equality.

Suppose Brenda is considering whether or not to smoke this cigarette, if she plays a prisoner's dilemma game with herself, the decision to smoke is always rational (assuming the present smoking will give her pleasure). Assume as in the prisoner's dilemma that Brenda does not know what her future selves will do with respect to quitting. This is not unrealistic. She presently reasons that if her future selves quit (do not defect), then there is no harm or loss in her now smoking this cigarette. Since she will soon be a non-smoker, thinking soritically and we surmise correctly, this cigarette will cost her nothing or at least nothing that will not soon be cancelled by her future healthy behavior. If on the other hand her future selves are non-quitters (they defect), then there is nothing to be gained by not smoking this cigarette. Since there will be future smoking anyway, not smoking this cigarette would be a sacrifice that would gain no reward. This latter reasoning holds even if some of her future selves opt not to smoke, as long as many opt to smoke. Thus the prospects for exposing weakness of will as irrational (in a purely self-interested sense) in a soritic context are bleak given standard game theoretic assumptions.

Appeal to moral-like reasoning involving the principle of equality is our best and perhaps only hope of exposing soritic thinking as irrational and thereby overcoming weakness of will. Direct appeal to purely self-interested rationality will not work. Even though the smoker and non-flosser are not involved in moral situations, moral-like

reasoning can be brought to bear in ways that can affect behavior. Of course, this appeal does not exclude the effectiveness of sanctions such as social scorn and so on. These have been particularly effective with helping people to quit smoking. However, in many cases of weakness of will involving soritic thinking such externalities will not be available. People are not scorned or punished for not flossing, or for eating another piece of cake. The most pressing externality is that the ‘collective’ good will not be accomplished—the goods of not getting lung disease, not losing one’s teeth, not being broke when one retires, not being overweight, etc. As we have seen, the problem is to connect the present micro-actions with achieving the overall ‘collective’ good. The most effective way is via universalizability and the principle of equality.

References

- Andreou, C. 2005: Going from bad (or not so bad) to worse, *American Philosophical Quarterly*, 42, 323-331.
- Andreou, C. 2006: Environmental damage and the puzzle of the self-torturer, *Philosophy & Public Affairs* 34, 95-108.
- Dummett, M. 1975: Wang’s paradox. *Synthese*, 30, 301-324.
- Eklund, M. 2005: What vagueness consists in. *Philosophical Studies*, 125, 27-60.
- Hampton, J. 1987: Free rider problems in the production of collective goods. *Economics and Philosophy*, 3, 245-273.
- Holton, R. 1999: Intention and weakness of will. *Journal of Philosophy*, 96, no. 5, 241-262 .
- Holton, R. 2004: Rational resolve. *The Philosophical Review*, 113, 507-535.

Schiffer, S. 2006: Vagueness. In Devitt, Michael and Hanley, Richard (eds.), *The Blackwell Guide to the Philosophy of Language*. Malden, MA: Blackwell Publishing.

Schwartz, S. 1989: Vagueness and incoherence: a reply to Burns. *Synthese*, 80, 395-406.

Sinnott-Armstrong, W. 2005: It's not my fault: Global warming and individual obligation. In W. Sinnott-Armstrong and R. Howarth (eds), *Perspectives on Climate Change: Science, Economics, Politics, Ethics*. Oxford: Elsevier.

Tsevat, J. et al 1991: Expected gains in life expectancy from various coronary heart disease risk factor modifications. *Circulation*, 83, 1194-1201.

Tuck, R. 1979: Is there a free-rider problem, and if so, what is it? In R. Harrison (ed.), *Rational Action*. Cambridge, Cambridge Univ. Press.

Unger, P. 1979a: There are no ordinary things. *Synthese*, 41, 117-154.

Unger, P. 1979b: Why there are no people. *Midwest Studies in Philosophy*, 4, 177-122.

Williamson, T. 1994: *Vagueness*. London: Routledge.

Wright, C. 1975: On the coherence of vague predicates. *Synthese*, 30, 325-365.

Wright, C. 1987: *Realism, Meaning, and Truth*. Basil Blackwell: Oxford.