Notes on Vagueness and Heuristics

Perceptual judgments often rely on various *heuristics*, fast and frugal ways of judging which are normally reliable, especially under ecologically realistic conditions, but are not 100% reliable. For example, we use discontinuities in colour as a heuristic in judging the boundaries of three-dimensional objects, which is mostly reliable but can be exploited by experts in camouflage to make us misjudge. Such ‘glitches’ in our heuristics cause predictable illusions in special circumstances. They are not individual performance errors, failures to execute the heuristic properly, or interferences with its outputs. Rather, they result systematically from its proper execution. We can regard the ability to apply the colour-discontinuity heuristic as a competence. Thus, in unfavourable circumstances, competent judgments can be false.

Paradoxes of vagueness may result from our reliance on useful but fallible heuristics.

When using a vague term, one can easily have the impression that some differences are too small to make any difference to its application. For example, a difference of one grain seems too small to make the difference between a heap and no heap. Thus we are tempted to assume that when one grain is removed from a heap of sand, leaving the rest undisturbed, what remains is still a heap. Such rules are known as *tolerance principles*: the vague term can tolerate small differences. Notoriously, tolerance principles generate sorites paradoxes, because many small differences can add up to a difference along the same dimension too large for the vague term to tolerate. Remove enough grains, one by one, and no heap is left.

 On some views of vagueness, competence with a vague term requires a disposition to find instances of a tolerance principle for it compelling. If so, I am not competent with any vague term, since I have no such disposition. Nevertheless, in everyday life, applying a tolerance principle is often reasonable. That suggests an understanding of tolerance principles as *heuristics* for applying vague terms, reliable but not perfectly reliable. They are cognitively efficient because they obviate the need for reassessing the application of the vague term from scratch to a case which differs only trivially from one already assessed. After recognizing a heap and removing one grain, you need not take another look at what is left to determine whether it still counts as a heap. You just apply the tolerance principle. When one unreflectively applies a tolerance principle in everyday life, the lurking threat of paradox need never occur to one.

 A complicating factor with vague terms is their tendency to shift in reference as the conversational context develops and new applications of them are charitably accommodated. However, such shifts are not inevitable. For example, when Mary tells John ‘Bring me a pot of red paint!’, what matters is the reference of ‘red’ in *Mary*’s context; for John to expand the reference of ‘red’ in supposedly obeying the instruction would be inappropriate.

 In what follows, the reference of the relevant vague terms will be taken to remain constant over the discourse. This is the most interesting case because the most resistant to tolerance principles, one-off instances of which are often easy to accommodate once contextual shifts are allowed. As noted in section 1.3, referential constancy is typically the appropriate assumption for discourse which integrates information originally linguistically encoded outside the immediate conversational context, at earlier times or by other people. Chapter 5 will discuss the issue of referential constancy in more depth and detail. In any case, as theorists of vagueness have observed, one cannot accumulate ascriptions of the vague term all the way along the sorites series. Sooner or later one comes to a shade one cannot count as ‘red’. Thus, even with contextual variation in reference, tolerance principles turn out to be fallible: they can postpone the moment of truth, but not indefinitely. To understand more clearly what is going on, we do best to focus on the simpler case without contextual variation.

 *How* reliable are tolerance principles? On predominant views of vagueness, when tolerance principles are formulated as universal generalizations, sorites paradoxes show them to be just false: they generate contradictions from clear cases of the application of the vague term and clear cases of its non-application. However, in everyday practice what we use is not the universal generalization but instances of a rule of inference. For a simple example, let us pretend that the application of ‘heap’ depends only on the number of grains. The false universal generalization is this (where the variable ‘*n*’ ranges over natural numbers):

HEAPU For every *n*, if *n* grains make a heap, *n* − 1 grains make a heap.

The rule of inference is this:

HEAP from: *n* grains make a heap

 to: *n* − 1 grains make a heap

A natural way to understand the reliability of an inference rule like HEAP is as the probability of the conclusion, conditional on the premise. For we want to know how good the rule is at preserving truth from premise to conclusion: in what proportion of cases where the premise holds does the conclusion also hold? By taking values of the variable ‘*n*’ as proxies for cases, we can reduce that question to the question: for what proportion of numbers *n* in the domain such that *n* grains make a heap do *n* – 1 grains also make a heap? Given the simplifying assumption that all those numbers are equally likely to arise, the answer to that question is just the conditional probability of the conclusion on the premise. For this purpose, instances of the rule with a false premise are irrelevant.

Some non-classical theories of vagueness undermine the distinction between truth and falsity so deeply that trying to estimate the reliability of a rule like HEAP has no obvious sense. To make progress, we will consider views on which there is a least number *k* of grains to make a heap, so *n* grains make a heap whenever *k* ≤ *n* and *n* grains do not make a heap whenever *n* < *k*. For example, on an epistemicist view of vagueness, there is a true exact specification of the value of ‘*k*’, but the vagueness of ‘heap’ prevents us from knowing what it is: many different exact specifications are consistent with what we know. Even more elusively, on a supervaluationist view of vagueness, usually treated as the leading rival to epistemicism, many different exact specifications of the exact value of ‘*k*’ are consistent with our vague meaning of ‘heap’; nevertheless, on each such specification, there is such a number *k*. On the former approach, the problem with ‘*k*’ is epistemic; on the latter, it is semantic, but both permit ordinary mathematical reasoning with ‘*k*’, even though it is vague what its value is.

Keeping the example realistic and mathematically straightforward, we assume that the domain of contextually relevant natural numbers is finite—say, all those less than a million—and treat all of them as equally probable values of the schematic letter ‘*n*’ in HEAP. In reality, there may be many more candidate-heaps with twenty grains than with twenty thousand, but for simplicity we can ignore such variations: taking them into account would not change the big picture. We also assume that the domain contains all relevant candidates to be *k*, the only value of ‘*n*’ for which HEAP has a true premise and false conclusion. By hypothesis, HEAP has a true premise for 1,000,000 – *k* values of ‘*n*’. Consequently, the reliability of HEAP is (999,999 – *k*)/(1,000,00 – *k*). For example, if *k* = 5,000, the reliability is 994,999/995,000. If *k* = 50, the reliability is 999,949/999,950. The latter value for ‘*k*’ is more plausible: fifty grains can be nicely heaped up. In short, HEAP is *almost* perfectly reliable. Only in the exceptional circumstances of a sorites series is HEAP liable to get one into trouble. As a heuristic for applying ‘heap’ in realistic circumstances, HEAP is excellent.

 Does vagueness in ‘*k*’ undermine such calculations of reliability? Even if we only know an upper bound on *k*, say *k* ≤ 10,000, we can calculate a lower bound on the reliability: it is at least 989,999/990,000, already very high. Any reasonable view of vagueness should make room for some such approximate knowledge articulated in vague terms, on pain of making vague language and thought useless in ways they obviously are not. For example, on an epistemicist account of vagueness, we can know that *k* ≤ 10,000 because it is true for every value of ‘*k*’ compatible with what we know. Similarly, on a supervaluationist account of vagueness, we can know that *k* ≤ 10,000 because it is true for every assignment of a value to ‘*k*’ consistent with its vague meaning, even though none of them is uniquely correct. Estimating the tolerance principle’s reliability in the way described does not commit one to any one view of vagueness.

 To vary the example, we can consider a case where the space of possibilities may be treated as continuous rather than discrete. Here is a tolerance principle for the vague term ‘about 6 o’clock’ (where the independent variables ‘*t*’ and ‘*t*\*’ range over times in a 12 hour cycle):

ABOUT6 From: *t* is about 6 o’clock

 *t*\* is within a minute of *t*

 To: *t*\* is about 6 o’clock

Suppose that the probability distribution over possible values of ‘*t*’ is uniform, so its value is equally likely to fall in any two periods of the same length. In the same spirit as before, suppose for the sake of argument that in the given context the times which qualify as ‘about 6 o’clock’ are just those between 5:55 and 6:05. Thus, given the first premise of ABOUT6, the probability of *t* being between 5:56 and 6:04 is 8/10; conditional on any such value and the second premise of ABOUT6, the probability of the conclusion is 1. Also given the first premise, the probability of *t* being between 6:04 and 6:05 is 1/10; conditional on any such value and the second premise, the probability of *t*\* being about 6 o’clock falls linearly from 1 to ½ as *t* goes from 6:04 to 6:05 (a time within a minute of 6:05 is just as likely to be before 6:05 as after it). Similar reasoning applies to times between 5:55 and 5:56. A standard probability calculation then shows that the reliability of ABOUT6 is (8/10 $×$ 1) + 2 $×$ (1/10 $×$ (1 + ½)/2) = 19/20. So ABOUT6 is very reliable too, though less spectacularly reliable than HEAP. On average, the conclusion is false one time out of twenty when the premises are true.

 We can easily ratchet up the tolerance principle’s reliability by narrowing the margin of tolerance in the second premise of ABOUT6, as a proportion of the period of times about 6 o’clock. For example:

ABOUTs6 From: *t* is about 6 o’clock

 *t*\* is within a second of *t*

 To *t*\* is about 6 o’clock

Formally, ABOUTs6 is just as capable as ABOUT6 of generating a sorites paradox, though the required sorites series of times will be longer. By a calculation just like the previous one, the reliability of ABOUTs6 is (598/600 $×$ 1) + 2 $×$ (1/600 $×$ (1 + ½)/2) = 1199/1200. On average, the conclusion of ABOUT6s is false less than one time out of a thousand when the premises are true.

 Although the numbers were obviously chosen with some arbitrariness, they are not unrepresentative. The calculations show clearly that tolerance principles can make highly reliable heuristics, even though the corresponding universal generalizations are just false. Moreover, by reducing the margin of tolerance in the second premise, we can make the reliability as high as we like, short of 1. No wonder we tend to rely on such principles.

 In many tolerance principles for supposedly observational terms, the second premise employs a standard of indiscriminability. Here is a standard example (where the independent variables ‘*x*’ and ‘*y*’ range over determinate shades of colour, and naked-eye indiscriminability in colour is meant):

RED From: *x* is red

 *y* is indiscriminable from *x*

 To: *y* is red

A reasonable conjecture is that the reliability of RED is not radically different from that of ABOUT6. It could be measured experimentally, given suitable operational tests for ‘red’ and ‘indiscriminable’. For example, one could consider a large set of determinate shades of colour, distributed more or less evenly over the colour sphere, and treated as equiprobable. The question is then: if you repeatedly pick *x* and *y* independently from the set, over a long series of trials, what proportion of cases verifying both premises also verify the conclusion?14

 When sorites paradoxes are presented as arguments, rules of inference such as HEAP, ABOUT6, and RED are often replaced by multiple instances of a corresponding conditional schema, treated like axioms:

HEAPif If *n* grains make a heap, *n* – 1 grains make a heap.

ABOUT6if If *t* is about 6 o’clock and *t*\* is within a minute of *t*, *t*\* is about 6 o’clock.

REDif If *x* is red and *y* is indiscriminable from *x*, *y* is red.

The argument is then driven forward by repeated applications of modus ponens, starting from a case where the vague term clearly applies. Each conditional can in turn be derived from the corresponding rule of inference, by an application of conditional proof. Of course, estimating the probability of such indicative conditionals in natural language depends on the semantics of ‘if’, which is at stake in this book.

Our immediate concern, however, is just to make the sorites paradoxes maximally challenging, for which purpose we need whatever reading of ‘if’ minimizes the strength, and so maximizes the probability, of the conditionals, while still validating modus ponens, without prejudice to the meaning of ‘if’ in English. That minimal reading of ‘if’ is the material one, so the conditionals become:

HEAP$⊃$ (*n* grains make a heap) $⊃$ (*n* – 1 grains make a heap)

ABOUT6$⊃$ (*t* is about 6 o’clock and *t*\* is within a minute of *t*) $⊃$ (*t*\* is about 6 o’clock)

RED$⊃$ (*x* is red and *y* is indiscriminable from *x*) $⊃$ (*y* is red)

As noted in section 3.2, Prob(*C*|*A*) $\leq $ Prob(*A* $⊃$ *C*) whenever the probabilities are defined, and indeed Prob(*C*|*A*) < Prob(*A* $⊃$ *C*) when in addition Prob(*A*) < 1 and Prob(*A* $⊃$ *C*) < 1, as would hold for at least some instances of the conditionals in a sorites paradox. Thus the material conditional versions of tolerance principles are even more reliable than the corresponding inference rule versions. Of course, the extra probability comes only from cases where the material conditional is true because its antecedent is false, and so does not concern the extra risk involved in moving from a true antecedent to the consequent. But the upshot is still that tolerance principles in material conditional form make at least as good heuristics as they do in inference rule form.

 In brief, sorites paradoxes are another example where philosophical paradoxes may have arisen because users of reliable but not perfectly reliable heuristics misinterpret and over-estimate their semantic status.

Arguably, *vagueness* also generates cases of non-transparent synonymy. On an epistemicist view, the vague adjective ‘small’, as applied to natural numbers, has the same truth-conditional meaning as the precise expression ‘at most *n*’, where ‘*n*’ stands in for the appropriate numeral (at least for a given context). But we are in no position to know exactly which numeral is the right one, because vague and precise expressions are associated with different cognitive practices. ‘At most *n*’ belongs to the language of pure arithmetic, and so is associated with the practice of applying various algorithms for numerical calculation, whereas the vagueness of ‘small’ prevents us from applying those algorithms to it. Instead, ‘small’ may be associated with a capacity which enables us to recognize some natural numbers (for instance, 0) as small, and others (for instance, in some contexts, a thousand) as not small, but fails to determine some intermediate cases either way.

As discussed above, many theorists have sought to understand vagueness and its tendency to generate sorites paradoxes in terms of *tolerance principles*, to the effect that if two things are similar enough in the respects relevant to a vague predicate, then the predicate applies to one of them only if it also applies to the other. Unfortunately, a finite sorites series starts with something to which the predicate recognizably applies and ends with something to which the predicate recognizably fails to apply, although any two neighbouring members of the series are as similar as the tolerance principle requires.

Despite the contradictions, tolerance principles are often claimed to be somehow built into the meanings of vague expressions, making those meanings inherently paradoxical. Competence with a vague expression may even be held to require a disposition to accept instances of the relevant tolerance principle. But even if speakers are so disposed, why regard the tolerance principles as anything more than fallible heuristics generally associated with vague expressions? As explained above, in any given context, almost all instances of the tolerance principle are truth-preserving inferences. The point is easiest to appreciate on a classical, bivalent semantics, which epistemicism about vagueness supports, but it applies at least as strongly on non-classical semantics for vague languages, since they are designed for the express purpose of making the tolerance principle fail less drastically than it does on the classical approach (at the cut-off point). Thus the tolerance principle is a *highly* though not *perfectly* reliable heuristic. That it gets us into trouble in a sorites paradox merely shows that with sufficient ingenuity one can exploit its limitations to make trouble, which is just what one would expect of a fallible heuristic.

Since tolerance principles are so simple and reliable for one-off applications, it is no wonder that we use them as heuristics for casual judgment. It would be surprising if we did not. A one-off applications of a tolerance principle does not draw attention to the limits of its validity. Its fallibility is not transparent to native speakers. We need paradoxes to warn us that something is wrong, and even then they do not tell us *what* is wrong. In using such heuristics pre-theoretically, we are not conscious of them as mere rules of thumb.

There is no reason why a tolerance principle for a vague term should be semantic. From the present perspective, tolerance principles are at the wrong level to figure in a compositional semantics for a vague language. Consequently, epistemicism about vagueness has no difficulty in giving tolerance principles their due at the level at which it is due. If speakers are disposed to use them, that is no evidence at all against epistemicism.

Indeed, making tolerance principles semantic add-ons undermines their role in explaining vagueness, because such add-ons are unnecessary for vagueness. For there could be a language just like English except that no tolerance principles had been added on. Removing the add-ons would not magically make the language precise. Tolerance principles are not special semantic rules. They are just manifestations of a more general and quite efficient cognitive tendency to treat apparently similar cases alike.