Toward a Distributed Computation Model of Extended Cognition[*]

Thomas W. Polger
University of Cincinnati
thomas.polger@uc.edu
http://homepages.uc.edu/~polgertw

1. The Pre-History of Extended Cognition

In the early years of the 1990s, a number of philosophers and cognitive scientists became enthused about the idea that mental states are spatially and temporally *distributed* in the brain, and that this has significant consequences for philosophy of mind.

Daniel Dennett (1991), for example, appealed to the spatial and temporal distribution of cognitive processes in the brain in order to argue that there is no unified place where or time when consciousness occurs in the brain. Dennett used this interim conclusion as a premise in a series of arguments designed to undermine what he calls the "Cartesian Theater" view of consciousness, according to which there is an independent and fact of the matter about when and where conscious mental states occur. Most famously, Dennett argues that anyone who believes there are such facts of the matter must choose between "Orwellian" and "Stalinesque" theories of the timing of consciousness, in order to accommodate certain data about the spatial distribution of the timing of consciousness-related brain events. But, he claims, that distinction is a distinction without a difference: The choice between "Orwellian" and "Stalinesque" views is either nonsense, or at the very least one that we could never be justified in making. And this, Dennett takes it, is a *reductio ad absurdum* of any theory of consciousness that says there is a fact of the matter about when and where is occurs.

[*] A version of this paper was presented at the Central Division of the American Philosophical Association in 2010. I am grateful to the audience on that occasion, and to Piotr Boltuc for the invitation to speak.

A second early advocate was Andy Clark (1996) who, inspired in part by roboticist Rodney Brooks (e.g., 1999), argued that much of our cognitive economy runs on piecemeal sensory and motor processes that are spatially distributed throughout the body, and is "scaffolded" by spatial and temporal relations that hold between the body and the world. Clark and Brooks took this psychological and engineering result to show that cognition does not require the kinds of unified and localized representational structures and cognitive modules that are the stock and trade of cognitive science, even in its connectionist forms. Like Dennett, Clark and Brooks hold that there is no single and total fact about the representational state of a cognitive system at any particular moment, or any single place where or time when the cognitive processing occurs.

In subsequent years, enthusiasm for these particular appeals to spatially and temporally distributed phenomena has faded. For one thing, despite a recent flurry of interest in the timing of the "feeling" of free will (Wegner 2002), it was widely appreciated that the timing phenomena cited by Dennett could be given alternative explanations.[1] And, with respect to the "non-representational" cognition suggested by Brooks, both he and Clark acknowledge that it would be more accurate to say that the systems lack central representations, rather than that they lack representations at all (Brooks 2002). So by the late 1990s there was sufficient reason to doubt that the spatial and temporal distribution of cognitive systems was as radical, or had as radical consequences , as its early proponents had claimed.

Furthermore, we might add, the very claim that cognitive systems, processes, or representations are spatially and temporally distributed was never subjected to careful

[1] See, e.g., Nahmias 2002.

analysis. It is not at all clear, for example, that paradigm cases of distributed systems, such as distributed computing systems, have the radical features that were supposed to go along with such distribution.

2. Embodied and Extended Cognition

Lack of enthusiasm for a few of the more radical arguments advanced by Dennett, Clark, or Brooks should not be understood as indicating that there is no interest in the spatially and temporally distributed processes that attracted their attention. Recently the focus has turned away from the ways that cognitive processes or states are spatially and temporally distributed within systems to the ways that cognitive states and processes may be spatially and temporally extended between systems, or between systems and their environments.

According to advocates of "embodied" cognition, including Clark and Brooks, cognitive processes constitutively involve "non-cognitive" aspects of the system in which (or of which) cognition occurs. These non-cognitive aspects might include the morphology of the system, as well as facts about the systems ability to interact with its environment that constrain cognition and action. Such constraints may modulate, simplify, and sometimes eliminate cognitive burdens that would have to be borne by the cognizing system according to traditional cognitive science. For example, a system might not have to calculate and initiate compensatory actions to maintain its balance, if instead the musculature of the system automatically makes such compensations.

And according to advocates of "extended" cognition, cognitive processes can constitutively involve distal objects and processes, or relations to them. That is, cognitive processes can constitutively involve processes that are external to the system's container, the "skin-bag" as Clark calls it in the case of human beings (2004). It is extended cognition that will be our topic henceforth, so it will be worthwhile to get clear about its core idea.

The canonical formulation comes from Clark's paper with David Chalmers, "The Extended Mind" (1998). There they argue that cognitive systems can become coupled with elements of their environments in ways that allow those external bits of the world to be actively involved in the cognitive processing itself—not just as objects or stimuli, but as genuine gears of the cognitive engine. Clark and Chalmers tell the story of two characters, Inga and Otto, who wish to find the Museum of Modern Art (MoMA) in New York City. Inga remembers that the MoMA is on 53rd Street, and she is able to use that memory to guide her action. Otto, however, suffers from a memory disability that does not allow him to recall the location of the MoMA in the same manner that Inga does. Fortunately, Otto is aware of his condition and plans accordingly: he always carries a notebook with him in which he writes down the information that he will need later, and he is able to quickly, reliably, and accurately make use of that information. So while Inga consults her memory, Otto consults his notebook.

But what is the difference between Inga's neurons and Otto's marks on paper? Both can be quickly, automatically, constantly, and reliably drawn upon to guide action. Indeed, Otto's notebook could be even more constant and reliable than Inga's memory:

> In these cases, the human organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right. All the components in the system play

> an active causal role, and they jointly govern behavior in the same sort of way that cognition usually does. If we remove the external components the system's behavioral competence will drop, just as it would if we removed part of its brain. Our thesis is that this sort of *coupled process* counts equally well as a cognitive process, whether or not it is wholly in the head. (Clark and Chalmers 1998: 8-9, emphasis added).

According to Clark and Chalmers, Otto's notebook can be coupled to him so that it is every bit as much a part of his cognitive system as Inga's memory-laden neurons are part of hers. The only difference is that Inga's cognitive processes are contained within her body, whereas Otto's cognitive processes "loop" into and are distributed between him and the world—they are extended cognitive processes (Figure 1).[2]

The extended cognition thesis has been widely elaborated since Clark and Chalmers introduced us to Inga and Otto. (And it had been widely criticized.) It has been frequently noted that the thesis depends on a broadly functionalist or computational theory of mind, and a rather abstract version no less.[3] And Fred Adams and Ken Aizawa argue that the extended cognition thesis depends on a fallacy, confusing causal coupling with constitution (e.g., 2001, 2008, 2009a, 2009b). Coupling, they argue, is perhaps necessary but not sufficient for constitution. For one thing, causal-functional coupling between cognitive/computational systems and t heir environments is abundant, indeed ubiquitous. Every creature is coupled with its environment in indefinitely many ways. But the thesis that cognitive processes extend outside the body and into the environment

[2] Here I am allowing myself to slide between talk of extended cognitive systems and extended cognitive processes, about which see Adams and Aizawa 2008. The distinction appears to be orthogonal to my present concerns.

[3] See, e.g., Shapiro 2007.

would be trivialized if every reliable causal interaction was a potential constitutive coupling. The interest of the thesis lies in the idea that there are actual but nevertheless relatively scarce kinds of cognitive coupling that allow extra-bodily processes to become constituents of cognitive processes. The question, then, is what scarce sorts of coupling are sufficient for cognitive constitution?

3. Back to Distribution

To answer the question of which sorts of coupling are sufficient for cognitive constitution, it will be useful to take a lesson from the early advocates of "distributed" cognitive systems. One trouble, recall, was that the fact that a system is spatially and temporally distributed does not by itself guarantee that there is no fact of the matter about what representational or computational state it is in at any particular moment. To put it in a slogan: Not all causal distribution is computational distribution. If cognition is a kind of computational or representational process, then it seems that spatially extending cognitive systems or processes is not sufficient for cognitively extending them. So if we're going to try to extend cognition by in some special way coupling spatially distributed components, not just any kind of coupling will do. The kinds of causal coupling that advocates of extended cognition need will be those that are cases of computational coupling. That is, we can offer them the view that physically coupled but spatially or temporally extended components are jointly constitutive of a cognitive system just in case the cognitive process they implement is itself computationally distributed/extended across them. What we need to find is systems that are not merely causally, spatially, and temporally distributed but also and *ipso facto* computationally distributed.

This is a nice proposal, but now we need to know what kinds of causal, spatial, and temporal distribution or extension are *ipso facto* computational distributions or extensions? Here's a suggestion: Let us suppose that a computational process will be distributed or extended if it operates on representations that are themselves distributed or extended. Why think this is a good approach? If the representations are not themselves extended, then it may be more plausible to say that the externally located representations are inputs or outputs of the system, rather that saying that they actively participate in the system. This is a natural response to the example of Otto's notebook: Representationally, Otto's notebook is fully self-contained. Despite Clark's and Chalmers' claims that the example illustrates "active" extended cognition (in contrast to the "passive" extension of Putnam/Burge-style wide content), the case is not so clear. Otto's notebook, though it is manipulated by Otto, is not manipulated by his cognitive system (a part of Otto) as a constituent of its control of Otto.[4] Rather, relative to Otto's cognitive system, the notebook is a passive piece of information or environment, a tool that can be taken up or not. It's representational value and content do not constantly

[4] This is something like the inverse of the "Systems Reply" to John Searle's "Chinese Room" example (1981). But in this case the complaint is that the notebook is not part of the cognitive system because it takes the whole cognitive system to manipulate it; no part of the cognitive system can make use of the notebook without all of Otto making use of the notebook. Not so with memories or beliefs, which can be used by parts of Otto (e.g. by Otto's associative learning systems) without Otto as a whole taking any action.

depend on the agent that make use of it, or on the fact that it is coupled to him.[5] The content is not itself representationally extended between the subject and the notebook.

A stronger case for extended cognition could be made if the cognitive system makes use of representations that are themselves constitutively extended into the external environment, and not fully contained either within or outside the system. Instead of cognitive processes that "loop" out into the world to pick up information, we would them have cognitive processes that make use of information that is already extended between the thinker and the world, not just available in the world to be accessed by the thinker (Figure 2).

At last we have come to the crux of the matter: What is it for a representation to be representationally distributed? We need a kind of representation that is not merely spatially or temporally distributed, but that is also and *ipso facto* representationally distributed.

4. Distribution and 'Real'

What is it for a representation to be representationally distributed? There are at least two good candidate answers to that question:

(i) Representations are distributed when their vehicle pieces *are not* individually semantically evaluable and syntactically structured.

(ii) Representations are distributed when their vehicle pieces *are* individually semantically evaluable and syntactically structured.

[5] Perhaps the representations in the notebook depend historically on being produced by someone; but that is not the kind of extended coupling that interests Chalmers and Clark, or other advocates of extended cognition.

How could these contradictory theses be options for understanding distributed representation?

Here some examples will be useful, so consider the representations of the number 173 that are illustrated in Figure 3. Most of these are standard representations in various numeral systems; let (h) be stipulated to be a representation of 173 for our purposes. Each of the examples (a)-(h) differ with respect to their spatial extension, and perhaps temporal extension as well. (c) is plainly more spatially distributed than (b) or (a), for example. And (a) and (h) appear to be most spatially compact, least distributed. (But if only the inked area is part of the representation proper, then (a) and several others may be smaller and less distributed representations than (h).) So Figure 3 illustrates a variety of representations of the same content that vary in their spatial and perhaps temporal distribution. Which of these representations is also (most) representationally distributed?

Right away we can notice that mere spatial and temporal distribution does not make for interesting representational distribution. (c) is more spatially extended than (b) and (a), and there is a plain sense in which its representational capacities are more distributed than theirs. Moreover, that distribution has practical consequences: some systems will take longer to process the information encoded in (c) than that in (a); using them may require different perceptual or storage systems, or at least the efficiency may be modulated by them. Yet it's fairly clear that the difference between (a), (b), and (c) is not what has usually interested philosophers and cognitive scientists. That difference, for example, in no way suggests that there is no determinate fact of the matter about what value is represented by (c), or that it is any less determinate than (a). We will return to this line of thought in a moment. But first let us turn our attention back to options (i) and (ii).

On the one hand, we might think that a representation is representationally distributed in case it's representational significance—its ability to function as a representation, and as a particular representation—is relatively immune to changes in its representational vehicle. In this case, example (g) seems to fit the bill. The prisoner's tally (unary) representation is robust in that it continues to be able to represent a quantity regardless of modifications to the vehicle, and the quantity or number represented approximates the original value in proportion to the modification to the vehicle. Not so for (a)-(c) where small changes to the vehicle can create large changes to the content. Or for the unstructured (h), whose content might be changed arbitrarily or eliminated by any changes to the vehicle. (Perhaps it is the case that a similar symbol that lacks the "leaf" element represents aardvarks.[6]) So there is a good sense in which (g) is representationally distributed, and indeed the most so of the examples in Figure 3.

On the other hand, every element within (g) has representational content independently of its participation in the whole representation. So the representational content of the spatially distributed (g) is merely a function of its composition out of elements whose representational capacity is (relatively) local and undistributed. In contrast, the representational capacity of (h) requires the presence and arrangement of all of its vehicle elements. And the representational capacity of (a)-(c) also requires the presence and organization of all the elements, or at least it is a non-summing function of the representational content of each element. (Otherwise (a)-(c) might all represent 11,

[6] We can stipulate that it is so. We can make similar stipulations about (g), but only by changing what or how (g) represents 173; that is, by changing a whole representational regime, not just stipulating about a single vehicle.

rather than 173.) Insofar as the ability of (a)-(c) or (h) to represent 173 depends on the presence of all of their vehicle elements and is not a summing function of the representational contents of those elements taken individually, there is a case to be made that the total representational capacity and content is distributed across all of the elements of the vehicle taken together.[7]

So there is a reasonable case to be made that either of the contradictory options, (i) and (ii), could be the right way to think about distributed representations. What are we to do? Here we can take a page from J. L. Austin. Though Austin was concerned with the ordinary meanings of words, his analytic tools can be adapted to say something useful about the technical concepts with which we are wrestling. In *Sense and Sensibilia* (1962) Austin argues that understanding the meaning of the word 'real' depends on a number of other factors, on what the subject is and what the speaker is trying to convey. So, for example, 'real' in "real duck" (real versus decoy) has a different significance than it does in "real oasis" (real versus mirage or hallucination.) Austin says that real is "substantive hungry" in that its meaning depends on the term (object) to which it is applied. Likewise, he think that 'real' depends on what it is contrasted with, so it is a contrast notion.[8] So the meaning of 'real' in "real duck" depends on the contrast being decoy ducks say, rather than robotic ducks. Sadly, we can readily imagine a future world in which someone says, "Those aren't real ducks, those are engineered. All the real ducks are in zoos." Austin also says of real that it is a dimension term, that it and its negation pick out the most general of a class. From more general to more specific:

---

[7] One might think that the elements must be coupled in a particular way. But then we will be running in circles.

[8] Austin says it is a "trouser" term.

unreal, artificial, decoy, wooden. Finally, he says that it is an adjuster term, that it allows us to say new things without introducing wholly new vocabulary, by adjusting the meaning of existing terms. We didn't have to introduce the term 'decoy' because we could readily say that some ducks are not real.

The suggestion, then, is that to answer the question of what kinds of representation are representationally distributed in space and time, we shall have to consider the ways in which distributed representations are supposed to modify and contrast with other kinds of representations. Plainly the notion of distribution here is used to adjust our standard notions of representation, the ones to which distributed representation—and, if I am right, distributed or "extended" cognition—are to contrast.

What about the substantive-hungry, contrast, and dimension aspects of distribution? Distributed representations are not just distributed ink splotches—this is why it seems right to say that (a)-(c) are not increasingly representationally distributed even if they are increasingly spatially distributed. This is why when the substantive whose distribution we are considering is representation—mere spatial or temporal extension is not sufficient. (Whether it is even necessary is unclear.) Likewise, we shall have to carefully attend to the kinds of representations to which distributed representations are supposed to contrast, and that lie on the axis of which distribution is one terminus.

At this point we know more about what kinds of representation are not distributed. Notice, in particular, that there is nothing special about the representations in Otto's notebook. They are not distributed: the representations in the notebook are fully encapsulated an local. This, I wager, is why the example of Otto's notebook is unconvincing to many philosophers and psychologists. It would be more convincing if Otto's notebook were only part of the representational vehicle—if the representations themselves were extended and non-local. Distributed representations are on one end of an axis of which local representations are the other extreme. This is the contrast to which distributed representations are compared, Of course the distinction is one of degree, but Otto's notebook seems to be on the wrong end of the spectrum.

5. Prospects for Extended Cognition

The reader will perhaps be disappointed, though not surprised, that I do not have a full theory of distributed representations to offer. Or even an outline of one. But that is not my goal. Rather, I have been concerned to sketch an alternative approach to extended cognition that does not depend on mere spatial and temporal extension. That, as we have seen, is at best necessary but not sufficient to produce extended computational systems. Actively extended computational systems—and thus actively extended cognitive systems, if cognition is a species of computation—should be systems that manipulate distributed and extended representations (per Figure 2). These will be representations that are already distributed between the core system and the world, not just those that the system can "loop" into the world to access (per Figure 1), however quickly or automatically.

I do not know whether there are any such systems in nature. And I have not argued that Clark, Chalmers, or someone else could not convince us that some "looping" systems are genuine examples of active extended cognition. But I suggest that the distributed representation model is worth exploring.

6. References

Adams, F. and K. Aizawa. 2001. The Bounds of Cognition. *Philosophical Psychology*, *14*: 43-64.

Adams, F. and K. Aizawa. 2008. *The Bounds of Cognition*. New York: Blackwell

Adams, F. and K. Aizawa. 2009a. Embodied cognition and the extended mind. In Garzon, P., & Symons, J. *Routledge Companion to the Philosophy of Psychology*: 193-213.

Adams, F. and K. Aizawa. 2009b. Why the Mind is Still in the Head. In *The Cambridge Handbook on Situated Cognition*: 78-95.

Austin, J. 1962. *Sense and Sensibilia*. Edited by G. Warnock. Oxford: Clarendon Press.

Brooks, R. 1999. *Cambrian Intelligence*. Cambridge, MA: MIT Press.

Chalmers, D. and A. Clark. 1998. The extended mind. *Analysis* 58:10-23.

Clark, A. 1996. *Being There: Putting brain, Body, and World Together Again*. Cambridge, MA; MIT Press.

Clark, A. 2004. *Natural Born Cyborgs*. Oxford: Oxford University Press.

Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown.

Dennett, D. and M. Kinsbourne. 1992 Time and the Observer. *Behavioral and Brain Sciences*, 15, 183-247.

Nahmias, E. 2002. When Consciousness Matters: A Critical Review of Daniel Wegner's *The Illusion of Conscious Will*. *Philosophical Psychology* 15(4): 527-541.

Searle, J. 1981. Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3 (3): 417–457.

Shapiro, L. 2007. The Embodied Cognition Research Programme. *Philosophy Compass* 2/2: 338–346.

Wegner, D. M. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.

Figure 1. The looping model of extended cognition.

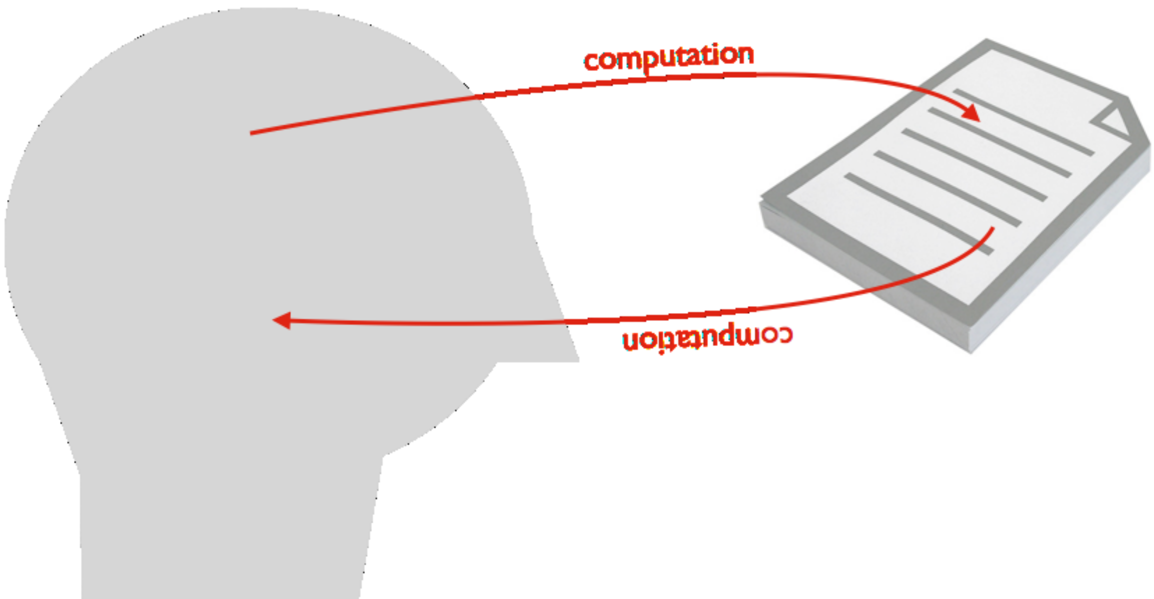


Figure 2. The distributed representation model of extended cognition.

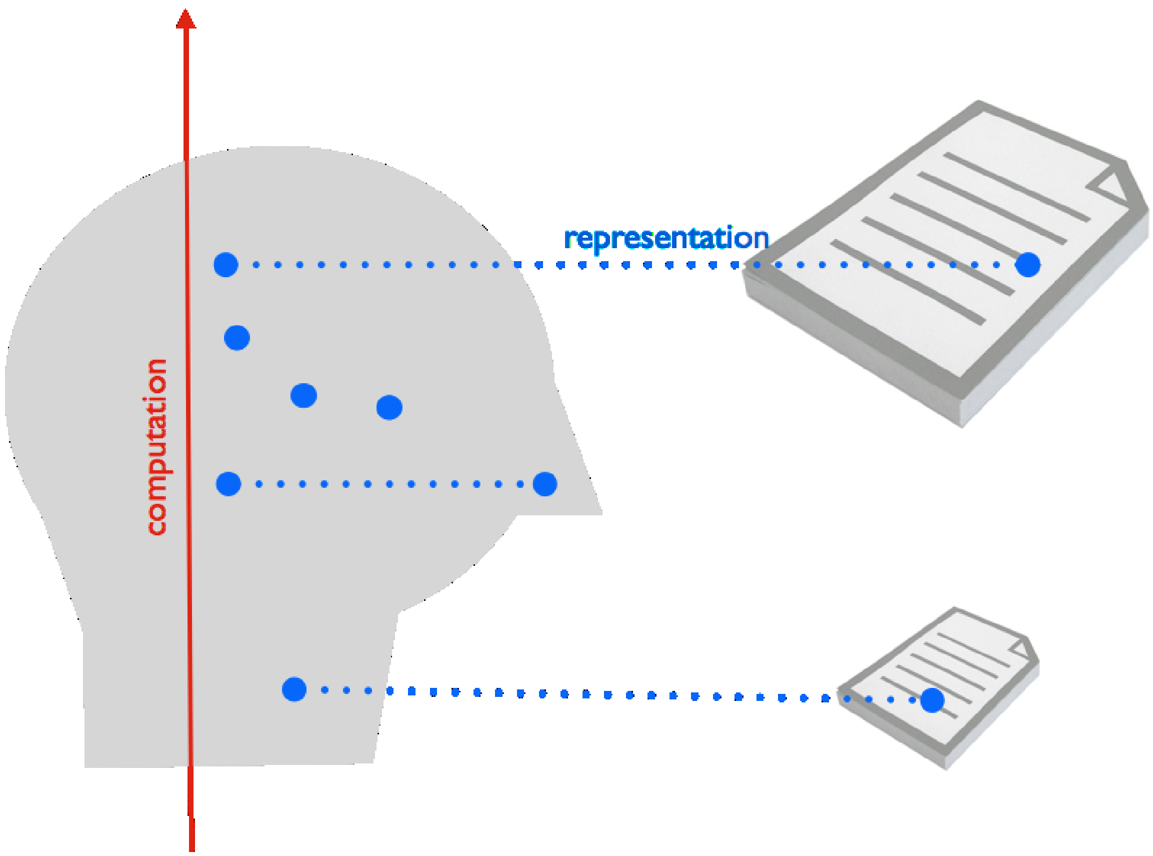


Figure 3: Distributed representations of the number 173.

(a)

173

(b)

173

(c)

173

(d)

one hundred seventy-three

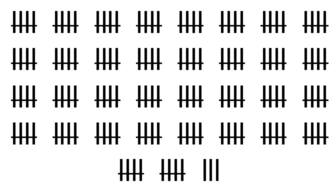
(e)

10101101

(f)

CLXXIII

(g)



(h)