	Theme		Rheme		comments
	textual or interpersonal Theme	topical Theme	expected Theme		
1.		attack evaluation			
2.		we		compare our targeted attacks to the best results previously in prior publication, for each of the three distance metrics	
3.		we		re-implement Deepfool, fast gradient sign, and iterative gradient sign	
4.		for fast gradient sign	we	search over ϵ to find the smallest distance that generates an adversarial example	Notice how Carlini and Wagner break with reader expectations both here and again at Position 10. This is well done, because at Position 3 the readers learns that three different methods are in discussion. That's a lot to handle all at once. However when a writer deftly manages the topical Theme as do Carlini and Wagner, then the three methods can get handled in a neat and orderly fashion.
5.		failure		is returned	
6.	if	no ϵ		produces the target class	
7.		our iterative gradient sign method		is similar	
8.		we		search over ϵ (fixing α = 1/256)	

CYBER SEC KIT GRADUATE SCHOOL

9.	and			return the smallest successful	This is a great way to introduce two equally important, but essentially separate points. Further examples are found at Positions 29, 31, 50, and 59.
10.		for JSMA	we	use the implementation CleverHans [35] with only slight modification (we improve performance by 50× with no impact on accuracy)	The parentheses de-emphasize content which otherwise would have provided more Theme and consequently, more topics. There is only one topic intended here: CleverHans was modified. You can see one further example of such use of parentheses at Position 77.
11.		JSMA		is unable to run on ImageNet due to an inherent significant computational cost	
12.		recall			See my discussion of such purely thematic clauses in the post for Part 5 of *Message in Text*. You can see further such examples at Positions 43 and 56.
13.	that	JSMA		performs search for a pair of pixels *p*,*q* that can be changed together that make the target class more likely and other classes less likely	This clause and the clause at Position 53 are only two examples in Section VII of poor construction in the grammar. The downranking here becomes just too complicated. The word *that* and then the word *that* again cause confusion in the Theme. Here is my understanding of the clause: JSMA performs search for a pair of pixels *p*,*q* which, when changeable together, will then make the target class more likely and other classes less likely.
14.		ImageNet		represents images as 299 × 299 × 3 vectors	
15.	so	searching over all pairs of pixels		would require 2^{36} work on each step of the calculation	

16.	if	we		remove the search over pairs of pixels	
17.		the success of JSMA		falls off dramatically	
18.		we		therefore report it as failing always on ImageNet	
19.		we		report success	
20.	if	the attack		produced an adversarial example with the correct target label	
21.	no matter	how much change		was required	
22.		failure		indicates the case where the attack was entirely unable to succeed	See my discussion of downranking in the post for Part 5 of *Message in Text*. You can see further examples of downranking at Positions 4, 13, 24, 35, and 78.
23.		we		evaluate on the first 1,000 images in the test set on CIFAR and MNSIT	
24.		on ImageNet	we	report on 1,000 images that were initially classified correctly by Inception v3	
25.		on ImageNet	we	approximate the best-case and worst-case results by choosing 100 target classes (10%) at random	
26.		the results		are found in Table IV for MNIST and CIFAR	

27.	and	Table V		for ImageNet	
28.		for each distance metric, across all three datasets	our attacks	find closer adversarial examples than the previous state-of-the-art attacks	An expert move putting into the Theme the conditions of the attacks, but into the Rheme the results of the attacks.
29.	and	our attacks		never fail to find an adversarial example	
30.		our L_0 and L_2 attacks		find adversarial examples with 2× to 10× lower distortion than the best previously published attacks	Notice how the Theme packages foregoing material. In Section III, we read phrases like *our attack for the* L_2 *distance metric*. But at this stage in the discourse, many sections on, all that information can easily be absorbed into the Theme as just *our* L_2 *attack*. This is one of the ways that Theme helps develop the topic of discourse.
31.	and			succeed with 100% probability	
32.		our L_∞ attacks		are comparable in quality to prior work	
33.	but	their success rate		is higher	
34.		our L_∞ attacks on ImageNet		are so successful	
35.	that	we		can change the classification of an image to any desired label by only flipping the lowest bit of each pixel, a change that would be impossible to detect visually	Here you see the optimal use of the Rheme — pack in there all the new information, one unit after another: Unit 1 *can change*, Unit 2 *the classification of an image*, Unit 3 *to any desired label*, Unit 4 *by only flipping the lowest bit of each pixel*, Unit 5 *a change that would be impossible to detect visually*. However, notice that Carlini and Wagner do not often pack the Rheme full in this way. Only six other Positions pack the Rheme to the same extent: Positions 10, 13, 45, 51, 68, and 78.

36.	as	the learning task		becomes increasingly more difficult	
37.		the previous attacks		produce worse results, due to the complexity of the model	
38.	in contrast	our attacks		perform even better	
39.	as	the task complexity		increases	
40.		we		have found	
41.		JSMA		is unable to find targeted L_0 adversarial examples on ImageNet	
42.	whereas	ours		is able to with 100% success	
43.		it is important to realize			
44.	that	the results between models		are not directly comparable	
45.	for example, even though	a L_0 adversary		must change 10 times as many pixels to switch an ImageNet classification compared to an MNIST classification	
46.		ImageNet		has 114× as many pixels	

47.	and so	the *fraction of pixels* that must change		is significantly smaller	Here is a unique example of downranking in the Theme, and it is well done This downranking inside of the Theme allows Carlini and Wagner to pack all of the relevant details about the pixels into one single topic, allowing them, in turn, to draw one single conclusion about that topic.
48.		generating synthetic digits			
49.		with our targeted adversary	we	can start from *any* image we want	Notice how Carlini and Wagner break with expectation in order to create expectation. The phrase *with our targeted adversary* is a break from the expectation of the word *we* as Theme. However, this break serves a purpose, namely, to shift the topic to Carlini and Wagner's own results. Moreover, the reader has now been primed to expect this construction when Carlini and Wagner's results will become the focus, and promptly, at Position 54 below, the authors use it again to the same effect. That is expert writing.
50.	and			find adversarial examples of each given target	
51.	using this	in Figure 6	we	show the minimum perturbation to an entirely-black image required to make it classify as each digit for each of the distance metrics	The phrase *using this* is really a textual Theme. It has, in this case, much the same meaning as *thus*.
52.		this experiment		was performed for the L_0 task previously [38]	

53.	however	when mounting their attack	one	for classes 0, 2, 3, and 5... can clearly recognize the target digit	My parse demonstrates how Carlini and Wagner have coerced the word-for-word quotation into this clause. I would edit the clauses at this and the previous Position to: This experiment was performed for the L_0 task by Papernot *et al.* [38]. Their attack demonstrates that for classes 0, 2, 3 and 5, the target digit is clearly recognizable.
54.		with our more powerful attacks	none of the digits	are recognizable	
55.		Figure 7		performs the same analysis starting from an all-white image	
56.		notice			
57.	that	the all-black image		requires no change to become a digit 1	
58.	because	it		is initially classified as a 1	
59.	and	the all-white image		requires no change to become an 8	
60.	because	the initial image		is already an 8	
61.		runtime analysis			
62.		we		believe	

63.		there		are two reasons why one may consider the runtime performance of adversarial example generation important	This thematic use of *there* has the function of presenting new and unexpected content. It is an attention-getter, because really, what this use of *there* means is, "Hey, do you know..." It is for this reason that I generally advise against overusing the phrases *there is* and *there are*. Note that in this entire section of the paper, Carlini and Wagner use it just this once. Their motivation is to present their reasons for calling the runtime performance important.
64.	first	to understand			
65.	if	the performance		would be prohibitive for an adversary to actually mount the attacks	
66.	second	to be used		as an inner loop in adversarial re-training [11]	
67.		comparing the exact runtime of attacks		can be misleading	
68.	for example	we		have parallelized the implementation of our L_2 adversary allowing it to run hundreds of attacks simultaneously on a GPU, increasing performance from 10× to 100×	
69.	however	we		did not parallelize our L_0 or L_∞ attacks	
70.	similarly	our implementation of fast gradient sign		is parallelized	

CYBER SEC KIT GRADUATE SCHOOL

71.	but	JSMA		is not	
72.		we		therefore refrain from giving exact performance numbers	
73.	because	we		believe	
74.		an unfair comparison		is worse than no comparison	
75.		all of our attacks and all previous attacks		are plenty efficient to be used by an adversary	Notice how here, winding down the section, Carlini and Wagner focus on the topic of attacks. Every Theme is related to attacks. And quite nice is how this and the next clause contrast like night and day: *all attacks — no attack*. That is clarity.
76.		no attack		takes longer than a few minutes to run on any given instance	
77.		when compared to L_0	our attacks	are $2\times-10\times$ slower than our optimized JSMA algorithm (and significantly faster than the un-optimized version)	
78.		our attacks		are typically $10\times-100\times$ slower than previous attacks for L_2 or L_∞, with the exception of iterative gradient sign which we are $10\times$ slower	I would replace the word *which* with either *where* or *in which*.

Commentary

I refer you to my commentary in the post for part 5 of *Message in Text*, because really, everything I say there applies here too.

It is notable, though, that in this section of experimental evaluation, Carlini and Wagner use the words *we* or *our* nearly forty times, whereas in Section III, a methodological section, they use only the word *we* eight times. In those word frequencies you have proof of how the grammar is there to realize the purposes of the authors. Every choice of word has a function, and in a section like Section VII, where the authors are running their own evaluations of the results, one absolutely crucial function of the grammar is to indicate which interpretations belong to whom.