	Given	New		comments
			Focus	
1.			attack algorithms	
2.			L-BFGS	
3.			Szegedy *et al.* [46] generated adversarial examples using box-constrained L-BFGS	Here you have proof that the systems of Given-New and Theme-Rheme are not just one system. This entire clause is parsed as New, but as you've seen in previous posts, this same clause is parsed into the Theme Szegedy et al. [46] and into the Rheme generated adversarial examples using box-constrained L-BFGS. Thus, the two systems do not align. Theme-Rheme is, if you will, created by the clause itself. Technically, the Theme-Rheme system is known as a structural system, which simply means that the system performs functions *inside of* the clausal grammar itself. And since the clause of English is very much a position-bound construct, Themes and Rhemes are likewise very much products of positions in the clause. Given-New is different. Given-New is what we might call supra-structural, because it is not a production of the clause, but instead the Given-New system is a production of the information *across* clauses. Really, the system of Given-New has its origins in our spoken intonation. There is a difference, for example, between "John wrote the ***pa***per" and "***John*** wrote the paper" and "John ***wrote*** the paper." Again, this is all quite involved, and really not immediately relevant to scientific prose, and my point anyway is not to explain precisely how Given-New has come about, but instead simply to demonstrate that the system comes about very differently to how Theme-Rheme does. So, to sum up, where the meaning of Theme-Rheme resides essentially in the positions which words take up in the clause, the meaning of Given-New resides in the sounds which words are given in spoken discourse, and those sounds are direct products of the Information the sounds convey to a listener. In this, you have the

				explanation for why intonation is so variable in the spoken language: the delivery of sounds is a major component of meaning-making in English. That being said, the two systems do *tend* to align, especially in scientific prose. Therefore, you can confidently search in the Rheme for the Focus of the New, as you can likewise and equally confidently search in the Theme for the Given. However, in the current clause, this wouldn't work, so again, it's good to note that the systems *are* essentially different.
4.	given an image x, their method	finds a different image x' that is similar to x under L_2 distance	yet is labeled differently by the classifier	From this clause onward, the system of Given-New begins to differentiate, and it's worth noting from the start here that there is necessarily some room for interpretation in the parses set out here in this table. The reason is this: Since the classes of Given and of New ultimately depend on the informedness of the reader, each reader will decide somewhat differently on just what to consider Given and just what to consider New. However, the leeway here is actually rather small, and smaller still in scientific prose, where authors create texts with very specific readers in mind. Nonetheless, this clause here certainly does offer some leeway. Here is how I read it, and let this serve also as my justification of parsing it as I have. The reader knows from clause No.3 that this portion of the text is about the generation of adversarial examples. The reader has also perused the Title and the Abstract and will certainly have scanned the first page; therefore, the reader knows that the experiments were conducted in the image domain. All this adds up to this understanding of the current clause: *So, they take some image — right, of course they do. And the method in question is this box-constrained L-BFGS. Yeah, and then there's this relation between* x *and* x' *— that I get. But, really, the big point here is the feature that distinguishes the* x'*. Ah, there it is, the label it receives. Okay, I see now.*
5.	they	model the problem as a constrained minimization problem	[mathematical problem]	

6.	this problem	can be	very difficult to solve	
7.	Szededy *et al.* solve	the following problem	[mathematical problem] where $loss_{F,l}$ is a function mapping an image to a positive real number	
8.	one common loss function to use	is	cross-entropy	
9.	line search	is performed to find	the constant $c > 0$ that yields an adversarial example of minimum distance	
10.	we	repeatedly solve this optimization problem for multiple values of c, adaptively updating c	using bisection search or any other method for one-dimensional optimization	
11.			fast gradient sign	
12.			the fast gradient sign [11] method has two key differences from the L-BFGS method	
13.	it	is optimized for	the L_∞ distance metric	
14.	it	is designed primarily to be fast	instead of producing very close adversarial examples	
15.	given an image x, the fast gradient sign method	sets	[mathematical problem] where ϵ is chosen to be sufficiently small so as to be undetectable, and t is the target label	

16.	for each pixel, the fast gradient sign method	uses the gradient of the loss function to determine in which direction the pixel's intensity should be changed (whether it should be increased or decreased)	to minimize the loss function	
17.	it	shifts all pixels	simultaneously	
18.	it	is	important to note	
19.	the fast gradient sign attack	was designed	to be *fast*, rather than optimal	Notice how punctuation serves to highlight the Focus. The italicized emphasis on the word *fast* tells you that this is focal information.
20.	it	is not meant to produce	the minimal adversarial perturbations	
21.			iterative gradient sign	
22.			Kurakin *et al.* introduce a simple refinement of the fast gradient sign method [26] where instead of taking a single step of size ϵ in the direction of the gradient-sign, multiple smaller steps α are taken, and the result is clipped by the same ϵ	
23.	begin	by setting	[mathematical problem]	
24.	on each iteration		[mathematical problem]	This is again one of those clauses with some leeway. In my reading, *an iteration* is considered to be, by this point in section III, to be a very familiar thing. However, it's possible you disagree and say that the word *each* really picks out a crucial procedural step which no reader

				could anticipate. If that is so — and again, it may well be so — then the phrase *on each iteration* belongs instead one box to the right.
25.	iterative gradient sign	was found to produce	superior results to fast gradient sign [26]	
26.			JSMA	
27.			Papernot *et al.* introduced an attack optimized under L_0 distance [38] known as the Jacobian-based Saliency Map Attack (JSMA)	
28.	we	give	a brief summary of their attack algorithm	
29.	for a complete description and motivation, we	encourage the reader to read	their original paper [38]	
30.	at a high level, the attack	is a greedy algorithm that picks pixels to modify one at a time	increasing the target classification on each iteration	Now, with this clause I want to recall the overarching aim of this series *Background on Text*: It is **to help you appreciate text so that you can utilize text for research purposes**. In pursuit of this aim, I have passed over and will continue passing over many matters in the linguistics which would only complicate text unnecessarily. For instance, it is an equally viable reading of this current clause here to say that *the attack* alone is the Given. The words *at a high level* would thus belong in the first box of the New (i.e., not in the Focus). Someone reading the sentence out loud would read with a marked rise in pitch at the word *high* then to fall again low in pitch by the first syllable of *level*. But that is just my point: The sentences of scientific prose are not spoken. Science is read silently. As a result, this distinction drawn by the intonation is lost. But on top

				of that, the distinction may even be too subtle, because the current reading as parsed here is just as viable as the alternative reading I am now entertaining for a spoken version of the sentence. Right, so again, please just know that the things I explain to you here are more complex than I make them out to be. And know too that I select what to explain and how far to explain it on the criterion of use-value to you.
31.	they	use the gradient [mathematical definition] to compute	a *saliency map*, which models the impact each pixel has on the resulting classification	The punctuation, again, announces the focal information, and so we read *saliency map* as a new term, and appended to the term, a quick in-context definition.
32.	a large value		indicates	
33.	changing it	will significantly increase	the likelihood of the model labeling the image as the target class *l*	
34.	given the saliency map, it	picks	the most important pixel	
35.	modifies it	to increase	the likelihood of class *l*	
36.	this		is repeated	
37.		more than a set threshold of pixels	are modified	
38.	which	makes the attack	detectable	
39.	it	succeeds	in changing the classification	
40.	in more detail, we	begin by defining the saliency map	in terms of a pair of pixels *p*,*q*	

41.	define		[mathematical definition]	
42.			α_{pq} represents	
43.	how much	changing both pixels *p* and *q* will change	the target classification	
44.			β_{pq} represents	
45.	how much	changing both pixels *p* and *q* will change	all other outputs	
46.	the algorithm	picks	[mathematical problem]	
47.	α_{pq}	> 0	(the target is more likely)	This and the next clause are fine examples of the difference between *generally* new information and *focally* new information. In both clauses, it is informative to learn about the relations between 0 and α or β, but the true significance of that relation is more informative still!
48.	β_{pq}	< 0	(the other classes become less likely)	
49.	$\alpha_{pq} \cdot \beta_{pq}$	is	largest	
50.			notice	
51.	JSMA	uses the output of the second-to-last layer *Z*, the logits	in the calculation of the gradient	
52.	the output of the softmax *F*		is *not* used	Once more, the punctuation tells the reader that this is Focus.
53.	we	refer to this	as the JSMA-Z attack	

54.	the authors	apply this attack to	their defensively distilled networks	
55.	they	modify	the attack	
56.	it	uses *F*	instead of *Z*	
57.	their computation	uses the output of the softmax (*F*)	instead of the logits (*Z*)	I argue for the current reading over this other possible reading. It is possible to say that the Focus is actually *the output of the softmax (*F*) instead of the logits (*Z*)*. But I disagree, because the connection between *logit* and *Z* is not strong. One clear function of this Focus, in fact, is to make that connection strong. If you look back to clause No.51, you will see the connection between *logit* and *Z* first being made. But that is not, at that place yet, a prominent piece of information. Therefore, I say that the prominence of the connection made here in the Focus of the current clause overrides the possibility of a long Focus (i.e., the entire phrase dependent on the noun *the output*).
58.	we	refer to this modification	as the JSMA-F attack	
59.	an image	has	multiple color channels (e.g., RGB)	The abbreviation *e.g.* really just elaborates on the class *multiple color channels*; in other words, one example of a *multiple color channel* is *RGB*. And since *RGB* is not exclusive (i.e., there will be other multiple color channels), I don't really consider it possible as a Focus on its own.
60.	this attack	considers the L_0 difference to be 1	for each color channel changed independently	
61.	all three color channels of one pixel change		change	

62.	the L_0 norm	would be	3	
63.	we		do not believe	
64.	this	is	a meaningful threat model	
65.	comparing to this attack, we	evaluate	under both models	
66.			Deepfool	
67.	Deepfool [34]	is	an untargeted attack technique optimized for the L_2 distance metric	
68.	it	is	efficient	
69.		produces	closer adversarial examples than the L-BFGS approach discussed earlier	
70.	the authors	construct Deepfool	by imagining	
71.	the neural networks	are totally linear	with a hyperplane separating each class from another	
72.	they	analytically derive	the optimal solution to this simplified problem	
73.		construct	the adversarial example	
74.	neural networks	are not actually	linear	
75.	they	take	a step towards that solution	

76.		repeat the process	a second time	
77.	the search		terminates	
78.	a true adversarial example	is	found	
79.	the exact formulation used	is	rather sophisticated	
80.	interested readers	should refer to	the original work [34]	

commentary

To really appreciate the system of Given-New, just run your eye once down the column Given, and then immediately after down the column Focus.

That impression is as true as the most precise functional-linguistic definition of either Focus or Given. In the column for Given you see a lot of words providing information that's both accepted and expected. Prime examples are the pronouns you read there.

The pronoun *it* occurs twelve times in this text, and only one occurrence is outside of the Given (in clause No.16). This is unsurprising when you consider that pronouns serve the function of referring back to what can already be presumed. By way of contrast, to find a pronoun in the Focus would surprise. This text, for example, does not have one instance of a pronoun in the Focus, and that one stray occurrence of *it* I mentioned is in the non-focal New — and there actually in parentheses. Parentheses generally have a sort of footnoting function, that is, parenthetical marks sequester off that portion of the line, both informationally and syntactically. This example of clause No.16 functions in just this way. First, notice how removal of the parentheses produces an ungrammatical sentence:

```
For each pixel, the fast gradient sign method uses the gradient of the loss function to determine in which
direction the pixel's intensity should be changed whether it should be increased or decreased to minimize the loss
function.
```

This tells us that we're dealing with a syntactical insert — the words *whether it should be increased or decreased* are, if you like, slid in here without regard to the normal progression of the other clauses in the sentence.

And the second reason for noting the parentheses here is that they affect, as well, the reference function of the pronoun *it*. Look again at the full sentence, and notice (a) the division into Given and New, and notice (b) the extent of the reference of *it* (boldfaced).

```
|Given For each pixel, the fast gradient sign method |New uses the gradient of the loss function to determine in
which direction the pixel's intensity should be changed (whether it should be increased or decreased) to minimize
the loss function.
```

That is a very short distance for reference, and really, little would change if we rephrased it thus:

```
|Given For each pixel, the fast gradient sign method |New uses the gradient of the loss function to determine in
which direction the pixel's intensity should be changed (whether the intensity should be increased or decreased)
to minimize the loss function.
```

In fact, in my view, the true motivation for writing it and not the intensity is simply the inclination to reduce material inside of parentheses. That's why, when inside of parentheses, writers opt for *e.g.* instead of the full *for example*; and it is my argument that this same inclination is in operation here in the material *whether it should be increased or decreased*. Basically, the reference-form of a pronoun is simply more compact than the reference-form of the reduced noun phrase *the intensity*.

Now, beyond the pronoun it, the text has instances of the pronouns *we* and *they*. The pronoun *we* occurs eight times in the text. All eight occurrences are in the Given. And the pronoun *they* occurs five times in the text. All five occurrences are also in the Given.

Again, these numbers just serve to illustrate the function of the Given to present the accepted and expected information, so that, all in all, the Given here meets reader expectations entirely.

The Focus, on the other hand, challenges reader expectations. You see in the Focus really the information which the reader does not yet know; however, you see in this information, too, exactly those things that a reader will normally be reading to find out. In short, the Focus contains the precise and specific information about results and about the interpretations of those results.

For example, here's clause No.54 — which I literally have selected at random:

|Given The authors |New apply this attack to their defensively distilled networks.

Actually, this clause is a bit of an exceptional case. Just note the use of the words which identify reference (boldfaced):

|Given **The** authors |New apply **this** attack to **their** defensively distilled networks.

It is, of course, not impossible to find such *determiners* as these in the New, but really, the expected position in the clause for identifying determiners is actually inside the Given. But here, in clause No.54, we find more determiners in the New than in the Given! The reason why is this: Clause No.54 is wrapping up the subheading *JSMA* and interpreting a good deal of the foregoing pieces of information in relation to one another. Basically, the new thing in this clause is the application of the attack we know, and the newer thing is what exactly this familiar attack is applied *to*. The fact that it's applied to a likewise familiar network is unimportant, because it's not the things themselves that are new (i.e., the attack or the network) but instead the way these things are related that counts as new.