Size of SAT Paradox

We noted that **3SAT** expressions for n variables have at most $O(\, n^3\,)$ clauses. Even including subscripts to identify the variables makes the expression length at most $O(\, n^3 \log n\,)$.

On the other hand there are 2^{2^n} Boolean functions of n Boolean variables. There are 2^n ways in which the n variables can be set to 0 or 1, and for each of these settings there are 2 choices for the output, 0 and 1. To represent these 2^{2^n} functions, at least 2^n bits are required, i.e. $\log \left[2^{2^n}\right]$.

PARADOX:
Since there are more functions than can be represented using $O(\, n^3 \log n\,)$ bits, **3SAT** expressions **cannot** represent all Boolean functions. BUT, every Boolean function can be put into **3SAT** form.

Resolution:
Consider a Boolean expression in clause form. The expression has k variables and C clauses. In general, all we know is that $C \leq 2^k$. To convert this expression to **3SAT** form, we take every clause with more that 3 literals and break it into 2 clauses by adding a new variable. E.G., if the clause is $(\, x_1 \vee \overline{x_2} \vee x_3 \vee x_4\,)$, we would replace this by $(\, x_1 \vee \overline{x_2} \vee y\,)$ and $(\, \overline{y} \vee x_3 \vee x_4\,)$. So if in a satisfying assignment x_1 were TRUE or x_2 were FALSE, we would choose y equals FALSE, and still have a satisfying assignment, since the original clause was TRUE and by our choice both of the new clauses are TRUE. On the other hand, if in a satisfying assignment x_3 or x_4 were TRUE, we'd choose y equals TRUE. and still have a satisfying assignment, since the original clause was TRUE and by our choice both of the new clauses are TRUE. Of course if x_1 and x_3 and x_4 were FALSE, and x_2 were TRUE, the original clause would be FALSE, and for either choice of TRUE or FALSE for y, one of the two new clauses would be FALSE.

By this construction, we will have added about C new variables (the y's), and our **3SAT** expression will have about $k + C$ variables. (A more exact analysis of the number of clauses is possible, but this is enough to resolve the paradox.) So, even though our **3SAT** expression will only have polynomial many clauses as a function of its variables, it has $k + C$ variables and this may be exponential in k the number of original variables.