Code Splitting Obfuscation - An Empirical Assessment

PoliTo,FBK

2/6/2017

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

Results

Descriptive statistics for the experiment

Table 1: Summary statistics for Correctness and Elapsed Time .

Treatment	n	n Correct	prop Correct	Time mean	Time sd	T normality
C	28	25	0.89	93.71	25.93	0
TS	29	15	0.52	99.28	27.84	0
T	30	20	0.67	93.07	29.19	0

H1 Correctness

The proportion of correct completed tasks for the three treatments is reported in figure 1

Table 2: Logistic regression of Correctness vs. Treatment

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.961	0.271	3.544	0.000
Treatment.L	-1.009	0.512	-1.973	0.049
Treatment.Q	1.092	0.423	2.580	0.010

Both splitting treatmens (T and TS) have a significant effect on the correcteness of the attack task results.

Looking at the coefficients we observe that when the reference treatment is applied – corresponding to the *(Intercept)* coefficient – we have a positive log-odds, i.e. the odds of a correct result vs. an incorrect one are larger than one. While the coefficients of the two treatments are neagitive, meaning that they lower the odds of a correct result.

H2: Time

The distribution of elapsed time required to complete the task is reported in figure 2.

Linear regression ANOVA

The Kruskal-Wallis test result (p-value = 0.3940098) does not allow us to reject the null hypothesis: neither obuscation treatment has a significant effect on the attack task time.

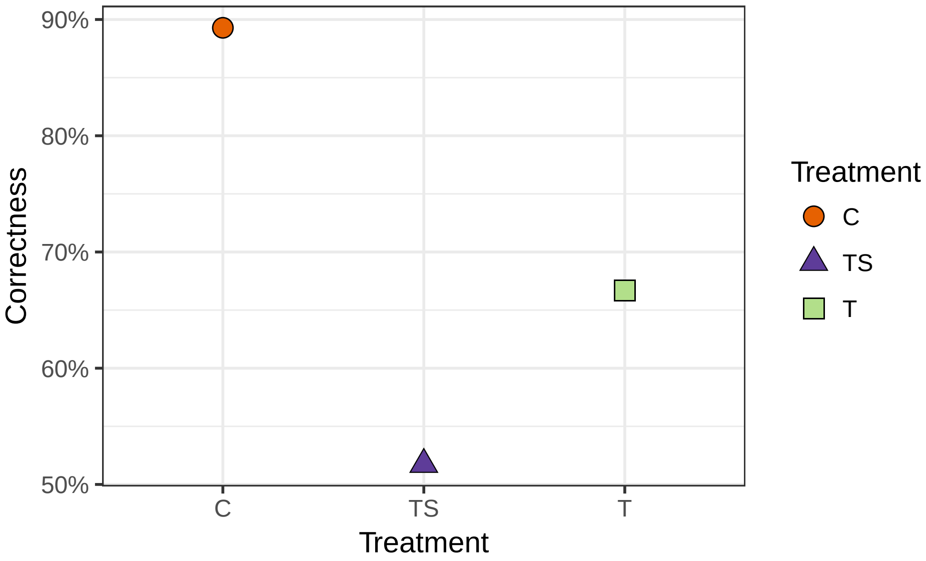

Figure 1: Proportion of correctly completed tasks per Treatment.

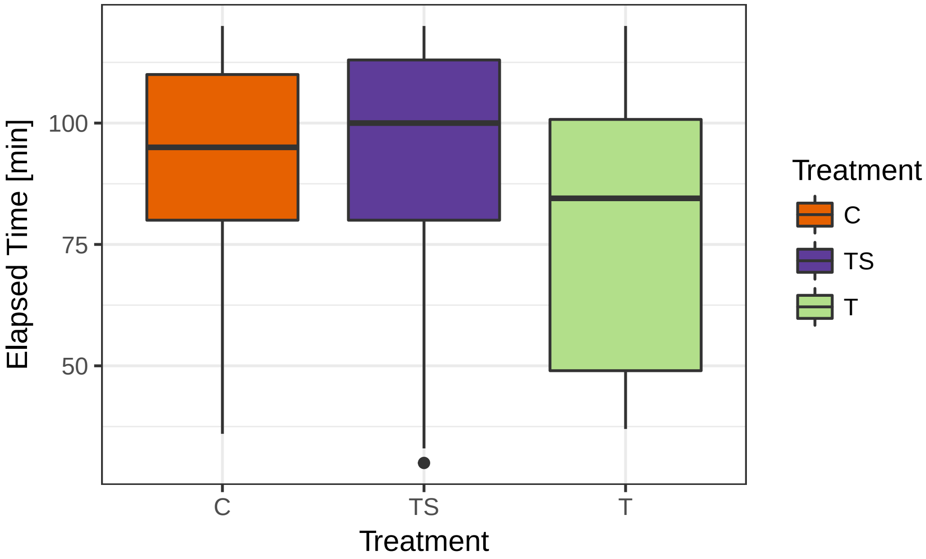

Figure 2: Boxplot of time to complete the task.

Co-factors

There three different tests for measuring the C knowledge skill. The cross correlation of the three tests results are shown in the table below:

Table 3: Cross correlation of the C knowledge tests

	Q1.TOT	Q2.TOT	Q3.TOT
Q1.TOT	1.00	0.49	0.13
Q2.TOT	0.49	1.00	0.35
Q3.TOT	0.13	0.35	1.00

Since they all the coefficients are relatively small, we decided to linearly combine the three tests outcomes to build a unique C knowledge score.

In addition we looked into the test 1 alone (just because we like fishing. . .).

Correctness and C Skill

The effect of the C knowledge co-factor is reported in figure 3 for both test 1 alone and the combined score.

The effect of the C knowledge, represented by the test 1 is reported in the following table:

Table 4: Logistic regression of Correctness vs. Treatment and C Test 1

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.756	1.191	-1.475	0.140
Treatment.L	-0.970	0.524	-1.853	0.064
Treatment.Q	1.120	0.442	2.533	0.011
Q1.TOT	4.153	1.808	2.297	0.022

The effect of the C knowledge, represented by the combined score is reported in the following tables:

Table 5: Logistic regression of Correctness vs. Treatment and Combined Score

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.941	1.345	-0.699	0.484
Treatment.L	-0.994	0.515	-1.930	0.054
Treatment.Q	1.109	0.430	2.579	0.010
C.SCORE	2.857	1.996	1.432	0.152

We observe no significant effect of the combine C knwoledge score on the correctness.

Time and C Skill

The relationship between the elapsed time and the C skill (both test 1 and combined score) is reported in figure 4.

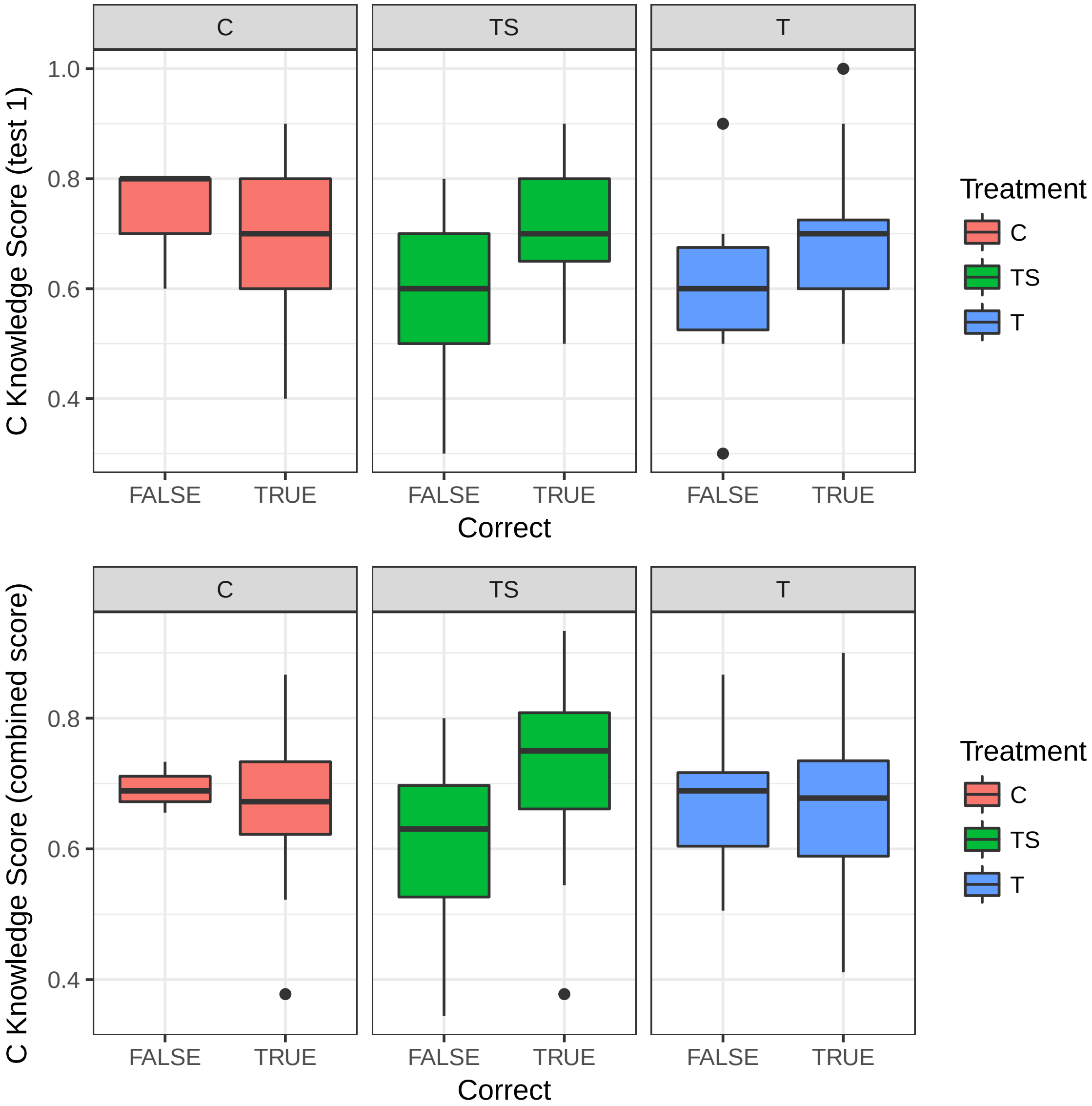

Figure 3: Proportion of correctly completed tasks by Treatment and C Score.

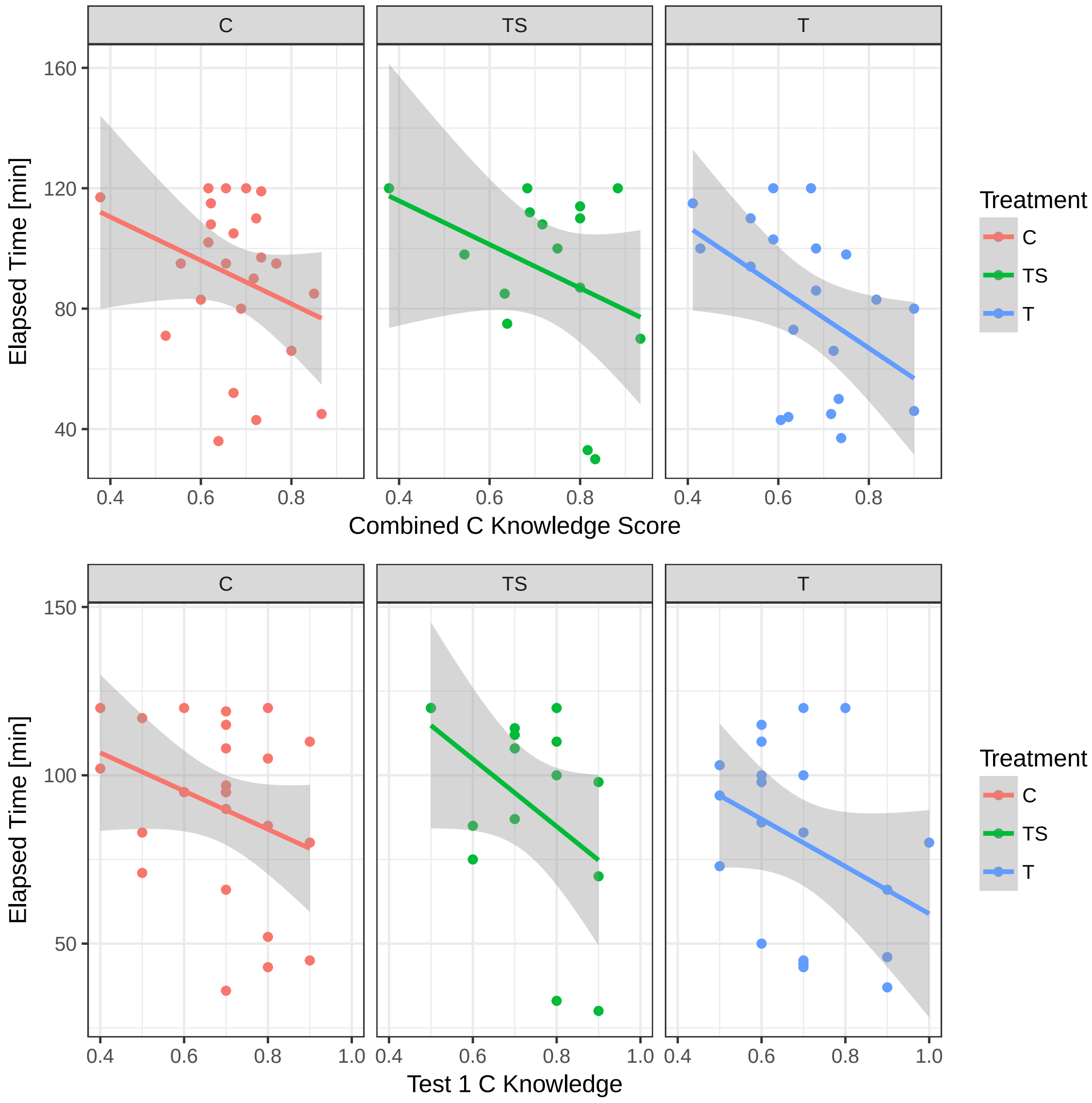

Figure 4: Boxplot of time to complete the task vs. Treatment and C skill.

The permutation test anova for C knowledge test 1 is reported in the following table:

Table 6: Logistic regression of Elapsed Time vs. Treatment and C Test 1

	Estimate	Iter	Pr(Prob)
(Intercept)	88.034	4598	0.021
Treatment.L	-6.706	241	0.295
Treatment.Q	-7.630	51	0.824
Q1.TOT	-70.994	5000	0.007

The test 1 C knowledge has a significant effect on the elapsed time. With test 1 we achieve an R^2 =0.1172919.

Table 7: Logistic regression of Elapsed Time vs. Treatment and Combined Score

	Estimate	Iter	Pr(Prob)
(Intercept)	88.134	3784	0.026
Treatment.L	-7.723	51	0.843
Treatment.Q	-9.198	342	0.228
C.SCORE	-83.152	5000	0.008

The combined score for C knowledge has a significant effect on the elapsed time.

With the combined score we achieve an $R^2 = 0.1228415$.

Type of attack

We identified two distinct categories of attack:

- internal: the computation inside specific functions has been altered,
- external: the actual parameter in function calls have been changed.

It is important to emphasize that during the attack task, both type could have been applied to achieve the goal.

Depending on the type of treatment one approach was easier to apply.

We can observe the number and percentage of participants that applied the two attach types in the following table.

Treatment	n	internal	internal.pct	external	external.pct	both	both.pct
C	28	20	71.4	13	46.4	5	17.9
TS	29	13	44.8	20	69.0	5	17.2
T	30	14	46.7	24	80.0	9	30.0

We observe that 2 participants did not reported which strategy was applied!

We observe that clear code was much easier to attack with an internal approach, the statistical significance of such a difference can be tested using a Fisher test comparing clear vs non-clear (either medium or small splitting).

##

```
##  Fisher's Exact Test for Count Data
##
## data:  d$Treatment == "C" and d$Internal
## p-value = 0.03757
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.032971 8.981842
## sample estimates:
## odds ratio
##   2.925874
```

The odds of an internal attack when a clear code is available are five times than for splitted code.

```
##
##  Fisher's Exact Test for Count Data
##
## data:  d$Treatment != "C" and d$External
## p-value = 0.01527
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.183089 9.683117
## sample estimates:
## odds ratio
##   3.332235
```

Concerning the external attacks, the odds of adopting such an approach are 2.5 times highen when a splitted code is present. Though such a difference is not statistically significant in our experiments. Several participants, even with clear code, did apply such a strategy.

Attack strategy vs. Success

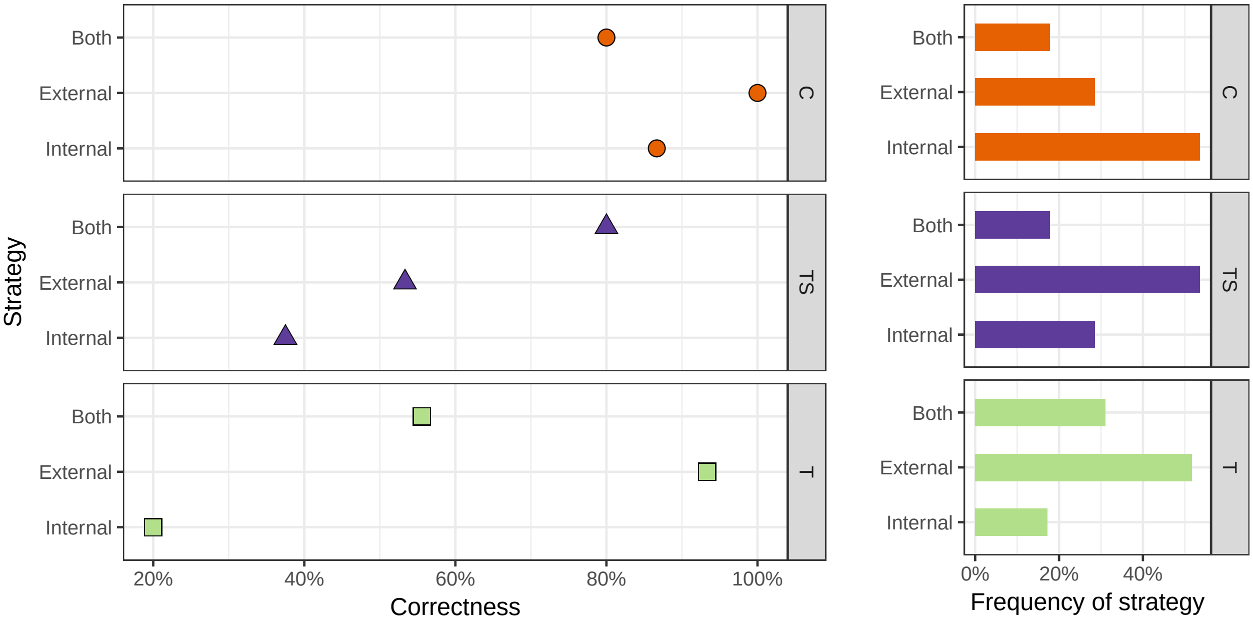

Figure 5: Success rate for different strategies by treatment

Attack strategy vs. Time

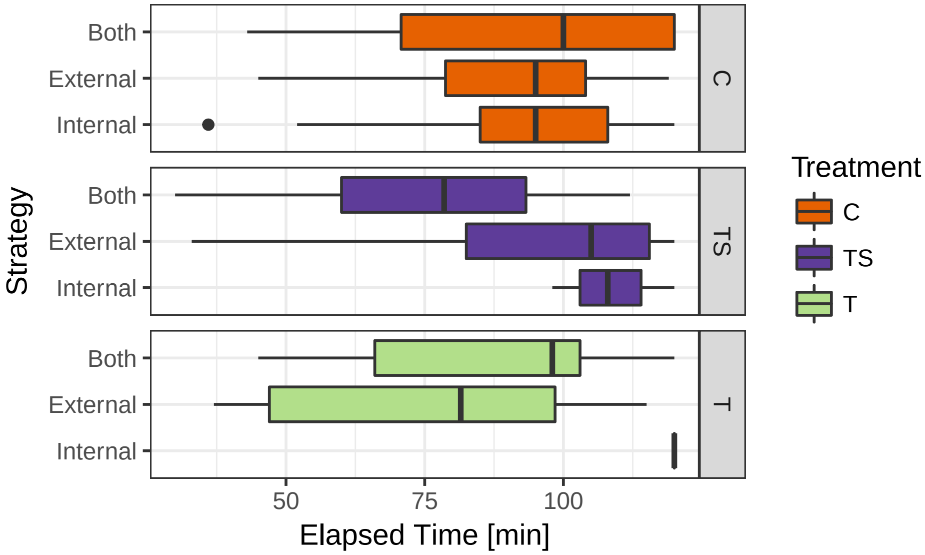

Figure 6: Attack time for different strategies by treatment (successfull attacks only)

Skill and strategy

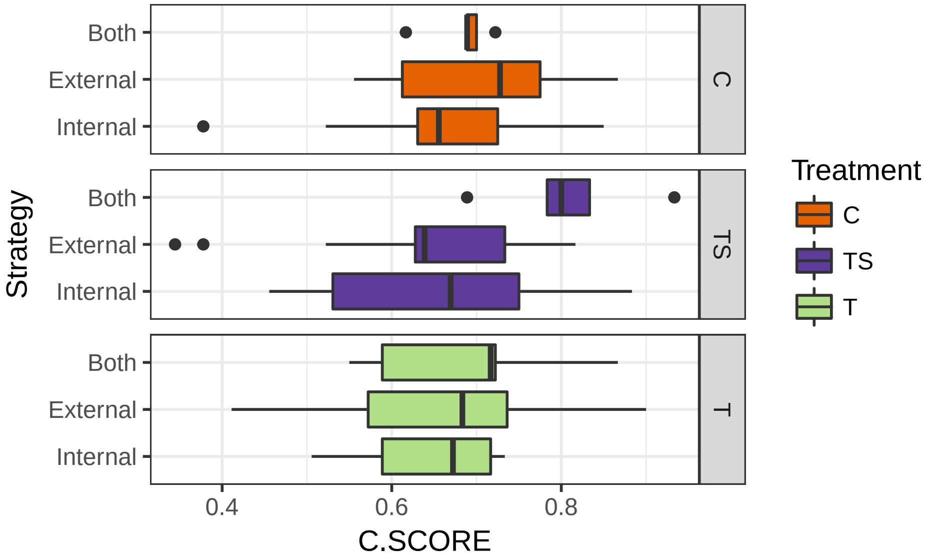

Figure 7: Skill of people adopting different strategies by treatment