Outcome and concepts: pilot study

Bartosz Janik

2023-04-05

Abstract

Introduction

The aim of this paper is to explore the possible connection between legal expertise and cognitive biases. It's basic motivation lays in (naive) belief that the legal concepts and the process of their application should at least partially reduce the susceptibility to cognitive biases.

Rachlinski and colleagues (2011) examined the decisions made by judges in cases in which the probable cause standard was used. They observed that when making those judgments judges might be prone to the hindsight bias as huge part of those judgments are made when determining whether to allow the admission of evidence obtained without a search warrant. The usual situation though is when the are determining ex ante whether to issue a search warrant. Based on the research on the hindsight bias they hypothesized that judges will not be able to set aside the acquired knowledge and judge probable cause as if they were working in foresight. They established, however that judges make similar rulings on probable cause in foresight and in hindsight. What is more, they have found that hindsight cloud judges' abilities to assess the likely outcome of the search (but only in certain cases), but does not influence their legal judgments in hindsight.

Tobia (2020) investigated the validity of reasoning from ordinary concepts to concepts of people with legal expertise. In his experiments he examined what it means to act intentionally for four populations of subjects: lay people, law students, non-law students and US judges (legal experts). His results indicate that legal training affected judgments. He further suggests, based on the data, that the phenomenon we are witnessing is an acquisition of a distinctive legal concept. The general conclusion of the study is that evidence about laypeople's ordinary concepts does not necessarily carry straightforward legal implications. What is interesting, Tobia's design (2020) in one of the experiments was Context (ordinary or legal) x Severity (moderate vs severe). In the Context condition the instruction for the participants was added indicating that the participant should imagine herself as a state judge in Dakota and relevant legal provisions were added.

Based on those two cases and bearing in mind other experiments from the experimental jurisprudence we tried to further extend their findings by checking to what extent categorical legal decisions and likelihood judgments might be prone to the outcome effect and if the information concerning definition and criteria of application of a legal concept might moderate the effect of an outcome. We thus assumed, following Tobia (2020) dissimilarity theory of legal concepts according to which the legal concept is similar to the ordinary concept in some ways, but is distinctive in other ways. What is more, we assumed that by asking the categorical question pertaining to the legal issue at hand, lay participants, having a required definition and criteria of application of legal concept will apply the concept correctly thus being immune to the outcome effect (Rachlinski, 2011).

Methods

Based of the theoretical insights, we hypothesized that even partial acquisition of the relevant concept should ameliorate the bias among the laypeople. This should be true bith in case of the categorical questions and when the Likert scale answer will be used. To test this hypothesis we constructed a simple vignette describing a legal case, similar to the case used by Rachlinski and colleagues (2011) but with certain changes to harmonize it with the polish criminal procedure (in polish criminal procedure a contept of "good reason",

similar to the concepts of a reasonable suspicion plays a role analogical to the role of probable cause). We decided to run this case only on laypeople. Our design was thus 2 Outcome (Neutral vs Bad) x 2 Legal information (Present vs Not present). After presenting the vignette, participants were asked whether a policeman had a good reason for conducting the search (Yes/No) and to assume a role of a prosecutor and to answer the question whether they approve the search (Yes/No). Third question asked to the participants was to assess the goodness of the reasons on a Likert scale (1-7).

Results

Sample

A total of 161 people were surveyed. Due to the legal background, seven people were excluded giving a final sample equal to 154. The mean age of the participants was 26.4, and the median age was 24. The majority of the sample was male (112(72.7%)), two people chose different gender, and one preferred not to answer this question.

Main analyses

Due to the binary nature of the first two questions (variables "pojazd_q0" and "pojazd_q2" respectively), a logistic regression analysis were conducted. Two separate logistic regressions, one for every categorical question, were performed with questions as response variables and outcome ("przeszukanie") and information ("info") as the predictor variables. The interaction term was also included in the model. What is more, the post hoc test pairwise comparisons of estimated marginal means using Bonferroni adjustment was performed.

```
library(readxl)
library(emmeans)
data <- read_excel("Outcome-expertise-PL-DATA.xlsx")
model1 <- glm(pojazd_q0 ~ przeszukanie * info, data = data, family=binomial)
summary(model1)
```

```
##
## Call:
## glm(formula = pojazd_q0 ~ przeszukanie * info, family = binomial,
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8394  -1.3153   0.6381   0.9587   1.0455
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.3185     0.3286   0.969    0.332
## przeszukanie         0.7802     0.4912   1.588    0.112
## info1                0.2205     0.4702   0.469    0.639
## przeszukanie:info1   0.1689     0.7277   0.232    0.816
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 189.48  on 153  degrees of freedom
## Residual deviance: 183.04  on 150  degrees of freedom
##   (1 obserwacja została skasowana z uwagi na braki w niej zawarte)
## AIC: 191.04
##
## Number of Fisher Scoring iterations: 4
```

```r
emms1 <- emmeans(model1, ~ przeszukanie * info)
pairs(emms1, adjust = "bonferroni")
```

```
##  contrast                                  estimate    SE  df z.ratio p.value
##  przeszukanie0 info0 - przeszukanie1 info0   -0.780 0.491 Inf  -1.588  0.6734
##  przeszukanie0 info0 - przeszukanie0 info1   -0.221 0.470 Inf  -0.469  1.0000
##  przeszukanie0 info0 - przeszukanie1 info1   -1.170 0.532 Inf  -2.198  0.1675
##  przeszukanie1 info0 - przeszukanie0 info1    0.560 0.496 Inf   1.127  1.0000
##  przeszukanie1 info0 - przeszukanie1 info1   -0.389 0.555 Inf  -0.701  1.0000
##  przeszukanie0 info1 - przeszukanie1 info1   -0.949 0.537 Inf  -1.768  0.4625
##
## Results are given on the log odds ratio (not the response) scale.
## P value adjustment: bonferroni method for 6 tests
```

```r
library(readxl)
library(emmeans)
data <- read_excel("Outcome-expertise-PL-DATA.xlsx")
model2 <- glm(pojazd_q2 ~ przeszukanie * info, data = data, family=binomial)
summary(model2)
```

```
##
## Call:
## glm(formula = pojazd_q2 ~ przeszukanie * info, family = binomial,
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7653  -1.2684   0.6876   1.0021   1.0891
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)           0.2113     0.3263   0.648    0.517
## przeszukanie          0.7581     0.4815   1.574    0.115
## info1                 0.2161     0.4654   0.464    0.642
## przeszukanie:info1    0.1362     0.7073   0.193    0.847
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 195.58  on 153  degrees of freedom
## Residual deviance: 189.41  on 150  degrees of freedom
##   (1 obserwacja została skasowana z uwagi na braki w niej zawarte)
## AIC: 197.41
##
## Number of Fisher Scoring iterations: 4
```

```r
emms1 <- emmeans(model2, ~ przeszukanie * info)
pairs(emms1, adjust = "bonferroni")
```

```
##  contrast                                  estimate    SE  df z.ratio p.value
##  przeszukanie0 info0 - przeszukanie1 info0   -0.758 0.481 Inf  -1.574  0.6923
##  przeszukanie0 info0 - przeszukanie0 info1   -0.216 0.465 Inf  -0.464  1.0000
##  przeszukanie0 info0 - przeszukanie1 info1   -1.110 0.515 Inf  -2.158  0.1856
##  przeszukanie1 info0 - przeszukanie0 info1    0.542 0.485 Inf   1.117  1.0000
##  przeszukanie1 info0 - przeszukanie1 info1   -0.352 0.533 Inf  -0.662  1.0000
##  przeszukanie0 info1 - przeszukanie1 info1   -0.894 0.518 Inf  -1.726  0.5061
```

```
##
## Results are given on the log odds ratio (not the response) scale.
## P value adjustment: bonferroni method for 6 tests
```

Overall, first model suggests that neither outcome nor info have a significant effect on the log odds of first question, and there is no significant interaction effect between these two independent variables. However, results, although insignificant, suggests that coefficient outcome has a positive effect on the log odds of the positive answer to the question.

Similarly, second model suggests that neither outcome nor info have a significant effect on the log odds of second question, and there is no significant interaction effect between these two independent variables. Similarly, results, although insignificant, suggests that coefficient outcome has a positive effect on the log odds of the answer to the question.

In both cases the post hoc pairwise comparisons of estimated marginal means using Bonferroni adjustment given no significant differences between the means.

For the third question ("pojazd_q3_1") a two-way ANOVA was performed to examine the effects of outcome and information. The post hoc HSD test was performed.

```
library(car)
data2 <- within(data, {
  przeszukanie <- factor(przeszukanie)
  info <- factor(info)
})
my_anova <- aov(pojazd_q3_1 ~ przeszukanie * info, data = data2)
hsd <- TukeyHSD(my_anova)
summary(my_anova)
```

```
##                     Df Sum Sq Mean Sq F value   Pr(>F)
## przeszukanie         1  37.80   37.80  20.053 1.48e-05 ***
## info                 1   0.83    0.83   0.441    0.508
## przeszukanie:info    1   0.01    0.01   0.005    0.946
## Residuals          150 282.72    1.88
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 obserwacja została skasowana z uwagi na braki w niej zawarte
```

```
print(hsd)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = pojazd_q3_1 ~ przeszukanie * info, data = data2)
##
## $przeszukanie
##          diff       lwr      upr    p adj
## 1-0 0.9908907 0.5536638 1.428118 1.48e-05
##
## $info
##           diff        lwr      upr     p adj
## 1-0 -0.1468779 -0.5841048 0.290349 0.5078582
##
## $`przeszukanie:info`
##               diff         lwr        upr     p adj
## 1:0-0:0  1.0039474  0.19595017  1.8119446 0.0082483
## 0:1-0:0 -0.1315789 -0.94986952  0.6867116 0.9753689
```

```
## 1:1-0:0  0.8421053  0.02381469  1.6603958 0.0410793
## 0:1-1:0 -1.1355263 -1.94352352 -0.3275291 0.0020215
## 1:1-1:0 -0.1618421 -0.96983931  0.6461551 0.9540790
## 1:1-0:1  0.9736842  0.15539363  1.7919748 0.0125761
```

The ANOVA output shows the results of testing the main effects of outcome and info, as well as their interaction (outcome:info). Based on it, there is a significant main effect of outcome on the assessment of the goodness of the reason, but no significant main effect of info or interaction between outcome and info. TheTukey HSD post hoc test result showed significant differences between conditions in which there were different outcomes. Eta squared and partial eta squared indicated that 11% of variance in the dependent variable is explained by the outcome which might be interpreted as medium or large effect (Cohen's f reported as well).

```r
library(effsize)
library(effectsize)
library(car)
library(lsr)

eta_sq <- etaSquared(my_anova)
print(eta_sq)
```

```
##                          eta.sq  eta.sq.part
## przeszukanie      1.171459e-01 1.175076e-01
## info              2.584544e-03 2.929124e-03
## przeszukanie:info 2.741745e-05 3.116315e-05
```

```r
cohens_f <- cohens_f(my_anova)
print(cohens_f)
```

```
## # Effect Size for ANOVA (Type I)
##
## Parameter         | Cohen's f (partial) |      95% CI
## -----------------------------------------------------
## przeszukanie      |                0.37 | [0.23, Inf]
## info              |                0.05 | [0.00, Inf]
## przeszukanie:info |            5.58e-03 | [0.00, Inf]
##
## - One-sided CIs: upper bound fixed at [Inf].
```

Exploratory analyses

```r
library(readxl)
data <- read_excel("Outcome-expertise-PL-DATA.xlsx")
model_exploratory1 <- glm(pojazd_q0 ~ przeszukanie * info + pojazd_q3_1, data = data, family=binomial)
summary(model_exploratory1)
```

```
##
## Call:
## glm(formula = pojazd_q0 ~ przeszukanie * info + pojazd_q3_1,
##     family = binomial, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3856  -0.3238   0.1980   0.4360   2.0772
##
## Coefficients:
```

```
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -6.34416    1.19346  -5.316 1.06e-07 ***
## przeszukanie        0.09408    0.73896   0.127    0.899
## info1               0.88743    0.70188   1.264    0.206
## pojazd_q3_1         1.71116    0.27160   6.300 2.97e-10 ***
## przeszukanie:info1 -0.39734    1.05523  -0.377    0.707
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 189.482  on 153  degrees of freedom
## Residual deviance:  99.274  on 149  degrees of freedom
##   (1 obserwacja została skasowana z uwagi na braki w niej zawarte)
## AIC: 109.27
##
## Number of Fisher Scoring iterations: 6
```

```
model_exploratory2 <- glm(pojazd_q2 ~ przeszukanie * info + pojazd_q3_1, data = data, family=binomial)
summary(model_exploratory2)
```

```
##
## Call:
## glm(formula = pojazd_q2 ~ przeszukanie * info + pojazd_q3_1,
##     family = binomial, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7470  -0.3725   0.2677   0.5467   2.0297
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -5.6699     1.0619  -5.339 9.33e-08 ***
## przeszukanie         0.0646     0.6793   0.095    0.924
## info1                0.7743     0.6488   1.193    0.233
## pojazd_q3_1          1.4861     0.2337   6.360 2.02e-10 ***
## przeszukanie:info1  -0.3358     0.9704  -0.346    0.729
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 195.58  on 153  degrees of freedom
## Residual deviance: 114.75  on 149  degrees of freedom
##   (1 obserwacja została skasowana z uwagi na braki w niej zawarte)
## AIC: 124.75
##
## Number of Fisher Scoring iterations: 5
```

```
data$above_thresh <- ifelse(data$pojazd_q3_1 > 3.7, 1, 0)
model_exploratory3 <- glm(pojazd_q2 ~ przeszukanie * info + pojazd_q3_1 + above_thresh, data = data, fa
summary(model_exploratory3)
```

```
##
## Call:
```

```
## glm(formula = pojazd_q2 ~ przeszukanie * info + pojazd_q3_1 +
##     above_thresh, family = binomial, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7693  -0.3687   0.2553   0.5426   2.0463
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -5.85250    1.38591  -4.223 2.41e-05 ***
## przeszukanie        0.06558    0.68000   0.096 0.923166
## info1               0.75812    0.65143   1.164 0.244515
## pojazd_q3_1         1.56612    0.45094   3.473 0.000515 ***
## above_thresh       -0.20222    0.96300  -0.210 0.833677
## przeszukanie:info1 -0.35311    0.97361  -0.363 0.716843
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 195.58  on 153  degrees of freedom
## Residual deviance: 114.71  on 148  degrees of freedom
##   (1 obserwacja została skasowana z uwagi na braki w niej zawarte)
## AIC: 126.71
##
## Number of Fisher Scoring iterations: 6
```

```
model_exploratory4 <- glm(pojazd_q2 ~ przeszukanie * info + above_thresh, data = data, family=binomial)
summary(model_exploratory4)
```

```
##
## Call:
## glm(formula = pojazd_q2 ~ przeszukanie * info + above_thresh,
##     family = binomial, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2283  -0.6358   0.4177   0.5565   1.9891
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -1.8293     0.5415  -3.378 0.000729 ***
## przeszukanie         0.4357     0.6096   0.715 0.474749
## info1                0.8646     0.6302   1.372 0.170052
## above_thresh         3.1804     0.4820   6.599 4.15e-11 ***
## przeszukanie:info1  -0.2561     0.9023  -0.284 0.776545
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 195.58  on 153  degrees of freedom
## Residual deviance: 131.19  on 149  degrees of freedom
##   (1 obserwacja została skasowana z uwagi na braki w niej zawarte)
## AIC: 141.19
```

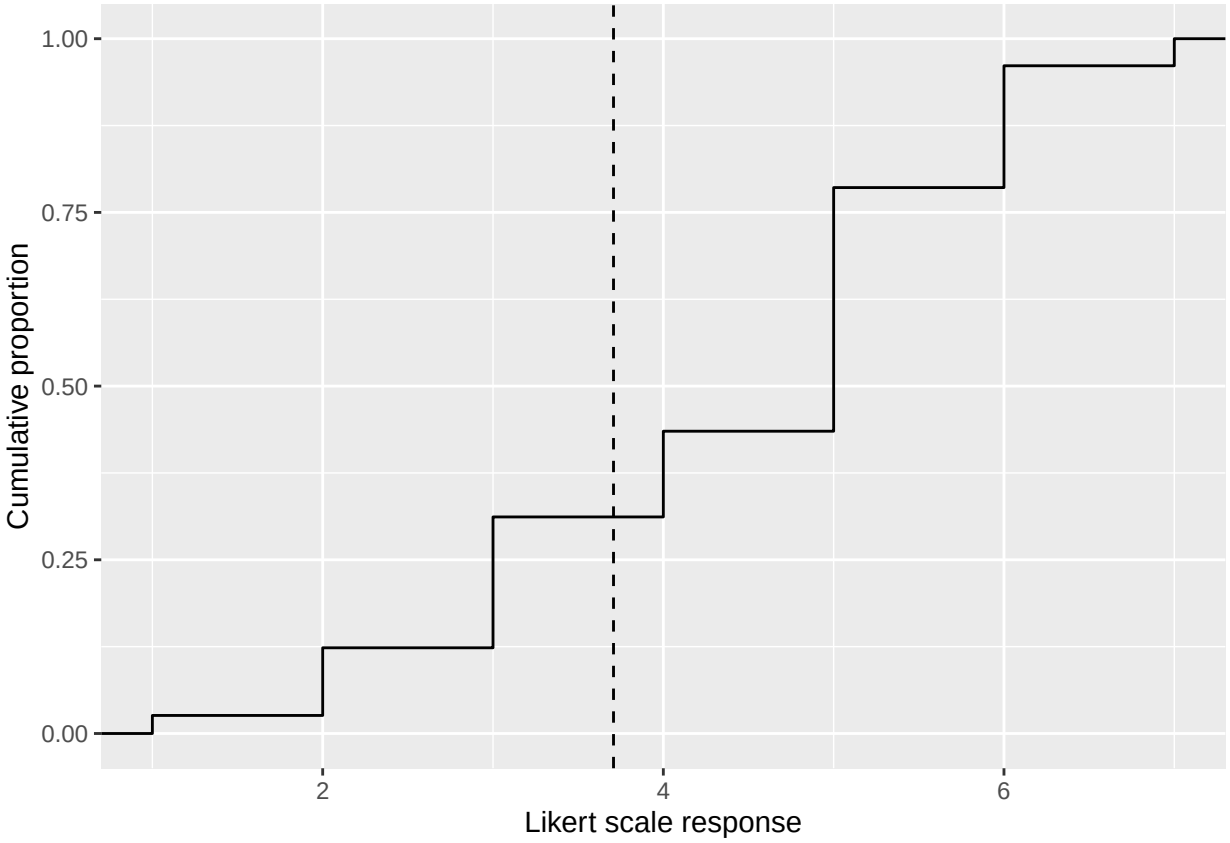
```
##
## Number of Fisher Scoring iterations: 4
```

```
threshold <- -coef(model_exploratory1)["(Intercept)"] / coef(model_exploratory1)["pojazd_q3_1"]
print(threshold)
```

```
## (Intercept)
##    3.707527
```

```
library(ggplot2)

ggplot(data, aes(x = pojazd_q3_1)) +
  stat_ecdf() +
  geom_vline(aes(xintercept = threshold), linetype = "dashed") +
  labs(x = "Likert scale response", y = "Cumulative proportion")
```



Discussion

This study is a mix of research programs on biases in legal decision making and on acquisition of legal concepts. Tobia (2020) claims that the second of those programs faces a distinctive problem as finding that the legal concept is different form an ordinary concept invites a discussion on to what extent they should be similar or dissimilar and the empirical study cannot offer the definite answer to that dilemma. This study aimes at investigation the problem on the intersection on this research programs, the possible interaction between legal information provided and susceptibility to the outcome effect in laypeople.

The results of the pilot experiment might be summarized in a few main themes.

1. For categorical questions outcome effect apparently is not working but for the continuous variable,

the main effect of outcome is significant. To account for that, we might assume that to answer the categorical question, one should have the appropriate gist (or a confidence level), which was not achieved in this experimental case. This might be further analyzed.

2. There was no effect of legal information. This lack of effect was observable for all three questions asked. This might be an effect of an insufficient exposition to the legal concept or lack of proper learning phase of the concept. The exposition to the information considering the content of the legal concept might be not enough to override the ordinary meaning of the concept that participants should possess. Further studies including the longer familiarization phase (or learning through examples) or using the concept that lacks the ordinary meaning should be conducted.
3. The results of the present study might be perceived as partially similar to the studies done before. For example this is a similar pattern to that observed by Rachlinski and colleagues (2011) because according to them legal judgments were not affected but the probability assessments were affected by the hindsight bias.
4. Based on Tobia's (2020) work and the results of the present experiment, one might ask to what extent the information given to the laypeople differed from the information associated with the folk concept of a "good reason". The possible explanation for the lack of an effect of information might be similar content of the folk and the legal concept of a good reason.

Limitations and further study

The main limitation of this study is the intuitive approach to concept acquisition in which we assumed that people presented with information concerning the concept are immediately gaining knowledge and (partial) expertise in concept application.

Moreover, the lack of a legal sample might be perceived as a limitation of the study and a significant obstacle to drawing inferences about the legal concepts and expertise.

Additionally, the use of a concept of a good reason, which might have strong folk meaning, might be problematic and possibly limit the scope of the possible conclusions.

Based on this, further research directions are:

1. Re-run this study on a professional sample.
2. Add additional protocols for concept acquisition and possibly ensure that subjects learn how to use the legal concept in question.
3. Test to what extent the results are replicated using different legal concepts, especially those which lack strong folk meaning (or their meaning is fully legal).
4. Test whether, by designing new cases, the threshold for a Yes/No legal decision might be different and could this might facilitate the outcome bias.

Appendix

Vignette

While on patrol in one of Krakow's neighborhoods, where a large rock concert was taking place, police officer Marcin P. noticed in the parking lot a nervously behaving man getting out of his BMW and rearranging some objects in the trunk of his car. The man then met up with an acquaintance, bought concert tickets from him and entered the hall where the concert was being held. After several minutes, the policeman noticed that one of teh windows in the car is half-open. Fearing that someone might commit a burglary, he approached the car. Upon approaching the car, he smelled something that seemed similar to the familiar smell of burnt methamphetamine from drug training a few years ago. The policeman looked into the car, but saw nothing that could be an illegal substance. Other than that, he noticed eye drops, a map of the area and several empty beer bottles in the car.

[Outcome-Neutral] Based on his experience and assessment of the situation, the policeman conducted a search of the car and found nothing in it.

[Outcome-Bad] Based on his experience and assessment of the situation, the police officer searched the car and found a package containing half a kilogram of methamphetamine hidden under the seat.

After the search, the police officer asked the prosecutor to approve the search and determine its legality.

[Info] Legal basis:

Article 219.

§1. A search may be conducted on premises and in other places to detect or arrest a person or to enforce their compulsory appearance, as well as to locate objects which might serve as evidence in a case or be subject to seizure in criminal proceedings, if there is good a reason to suppose that the suspected person or the objects are to be located there.

§2. In order to find the objects referred to in § 1 and on the condition set forth in that provision, and having regard to the rules and limitations set out in Article 227, a search of a person, their clothing and objects at hand may also be carried out.

Comment: according to the literature, the supposition in question is a subjective judgment about what was, is or will be in reality, arising in the psyche of a person under the influence of the information he has about events, knowledge of the regularities in the surrounding world, individual ability to associate facts and other factors. Moreover, the decision to search must have a rational justification and be based on a real probability derived from the reliable information gathered; it is not permissible to intuitively search any place with which the wanted person has ever had contact.

Questions

Q1: At the time of the decision, did the police officer have reasonable grounds to suppose that drugs might be in the car and could have searched the car? [Yes/No]

Q2: Imagine that you are a prosecutor who has to make a search approval. The decision hinges on the recognition that the police officer had reasonable grounds to believe that there were drugs in the car. In your opinion, at the time of the decision, did the police officer have reasonable grounds and you approve the search? [Yes/No]

Q3: How justified was the police officer's basis for supposing that there were drugs in the car? [Likert scale: 1-7]