ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

UNIVERSITÄT BERN

Sensitive Question Techniques in Online Surveys: An Experimental Comparison of Different Approaches

Andreas Diekmann (ETH Zurich)
Marc Höglinger (ETH Zurich)
Ben Jann (University of Bern)

Rational Choice Seminar, Venice, December 1st, 2011

The problem

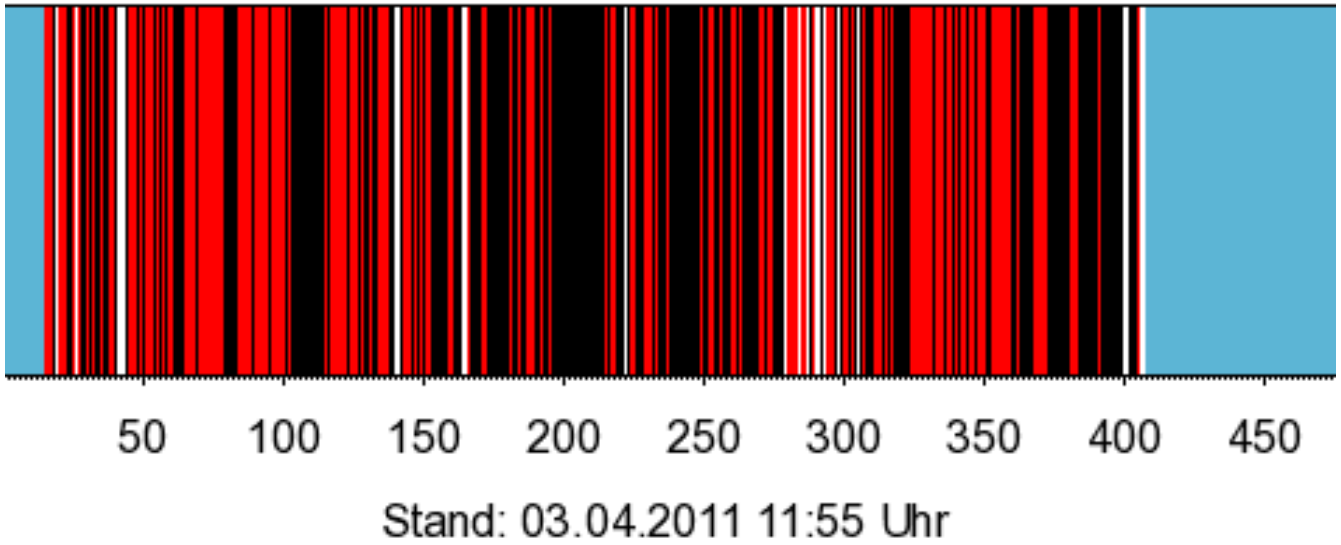
Conventional questioning in surveys, so-called direct questioning (DQ), does often not work out well when asking sensitive questions…

…for instance on scientific misconduct such as plagiarism.


ZDF

1218 plagiarized passages from 135 sources on 371 out of 393 pages (94.4%)

source: http://de.guttenplag.wikia.com



- pages with plagiarized passages
- pages with plagiarized passages from various sources
- pages where no plagiarized passages have been found
- table of content and appendix

Outline

- Sensitive questions in survey research
- Some indirect approaches to elicit truthful answers
  - The Randomized Response Technique (RRT)
  - The Crosswise Model (CM): A new alternative to RRT
- Experimental comparison of the different approaches: a survey on students' cheating and plagiarism
- Conclusion

Eliciting truthful answers to sensitive questions – not an easy task

- Survey respondents might not tell the truth if asked questions on sensitive issues. This leads to distorted results.
- Examples for proportion of ‘liars’ (respondents with a false negative response) in surveys that use direct questioning (estimates from validation studies):
  - Penal conviction: 42.5% (F2F, Wolter 2010)
  - Welfare and unemployment benefit fraud: 75% (F2F, van der Heijden et al. 2000)
  - Driving under influence: 54% (P&P, Locander et al. 1976)
  - Bankruptcy: 32% (Ibid.)

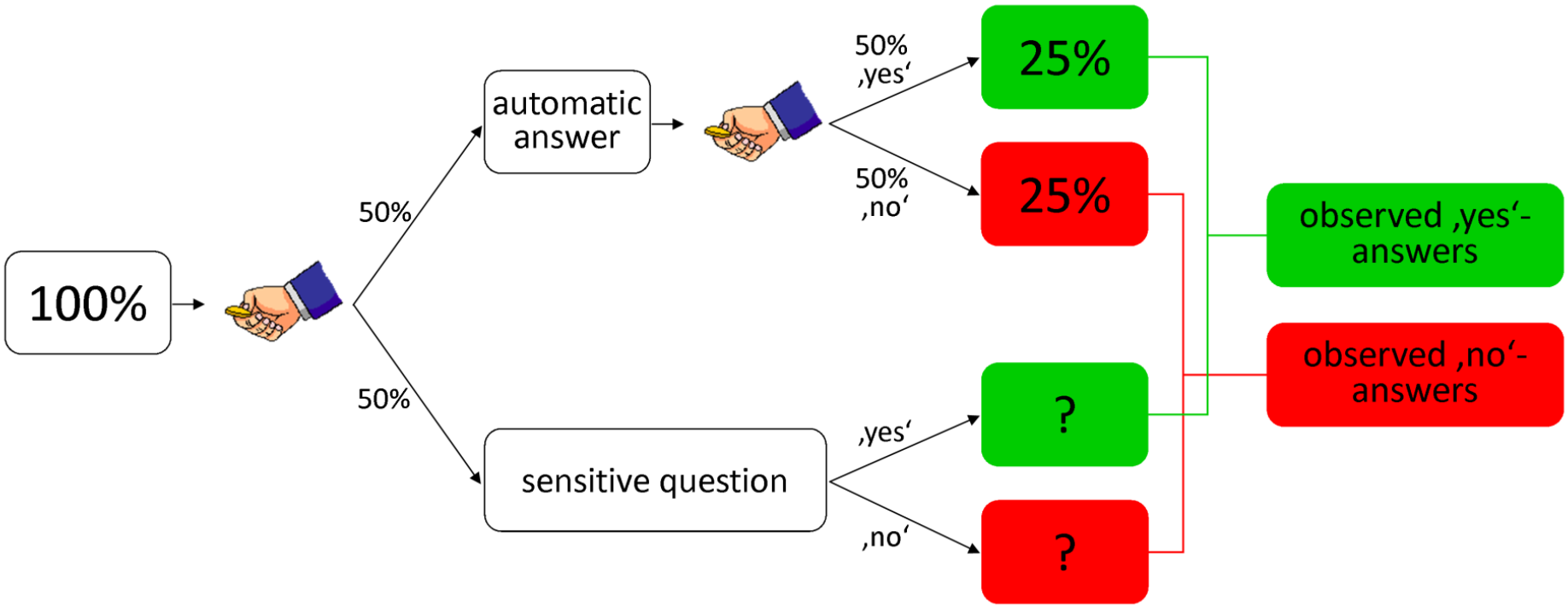
The Randomized Response Technique (RRT)

(Warner 1965; Fox and Tracy 1986)

- Main principle: privacy protection through randomization (i.e. add random noise to the answers)
- A randomizing device, the outcome of which is only known to the respondent, decides whether…
  - the sensitive question has to be answered
  - an automatic ‘yes’ or ‘no’ has to be given or a surrogate question has to be answered
- Since only the respondents knows the outcome of the randomization device, a ‘yes’ cannot be interpreted as an admission of guilt.
- However, with knowledge of the probability properties of the randomizing device, a prevalence estimate for the sensitive question can be derived.

example

RRT Example (forced response version)



- Prevalence estimate:

$$Pr(observed\ yes) = Pr(sensitive\ question) * \pi + Pr(automatic\ yes)$$

$$\pi = \frac{Pr(observed\ yes) - P(forced\ yes)}{Pr(sensitive\ question)}$$

example

The Crosswise Model (CM): A new alternative to RRT

(Yu, Tian, and Tang 2008)

- Simple idea: Ask a sensitive question and a nonsensitive question and let the respondent indicate whether …
  - answers to the questions are the same (both ‘yes’ or both ‘no’)
  - answers to the questions are different (one ‘yes’, the other ‘no’)

		nonsensitive question	
		No	Yes
sensitive question	No	same	different
	Yes	different	same

- Note: Questions must be uncorrelated and probability of ‘yes’ must be unequal 0.5 for the nonsensitive question.

example

The Crosswise Model (CM): A new alternative to RRT

(Yu, Tian, and Tang 2008)

- Prevalence estimate:

$$Pr(same) = (1-\pi) * (1 - \Pr(nonsensitive\ yes)) + \pi * \Pr(nonsensitive\ yes)$$

$$\pi = \frac{Pr(same) + Pr(nonsensitive\ yes) - 1}{2 * Pr(nonsensitive\ yes) - 1}$$

- Note: CM is formally identical to Warner's original RRT model

example

Performance of RRT and Crosswise

- RRT does not seem to work well in in online surveys.
  - no different prevalence estimates than with direct questioning (Coutts & Jann 2011, Peeters 2006, Snijders & Weesie 2008)
  - lower prevalence estimates than with direct questioning or even negative estimates (Coutts et al. 2011, Holbrook & Krosnick 2010, Coutts & Jann 2011)
- However, RRT implementations so far were often not well suited to survey mode.
  - random device not at respondents' immediate reach
  - random device not trustworthy

Performance of RRT and Crosswise

- The Crosswise Model seems to work better
  - higher prevalence estimates than with direct questioning in a p&p survey on plagiarism (Jann, Jerke, Krumpal 2011)
  - however, no empirical application in online mode so far
- Advantages of the Crosswise Model over RRT
  - easier to understand
  - no need for a randomizing device
  - respondent is not forced into giving a ‘false’ automatic response or a seemingly irrelevant response to an innocuous instead of a sensitive question
  - no obvious self-protective answering strategy (e.g. always say ‘no’)

The study: survey on student cheating and plagiarism

- Web survey among students of University of Bern and ETH Zurich in spring 2011
- Response rate 33%, N=6‘485
- Comparing direct questioning (DQ) to three variants of RRT and two variants of the Crosswise Model (CM)
- Sensitive questions on
  - copying from other students in exam (copy)
  - using crib notes in exam (notes)
  - taking drugs to enhance exam performance (drugs)
  - partial paper plagiarism (partial)
  - severe paper plagiarism (severe)

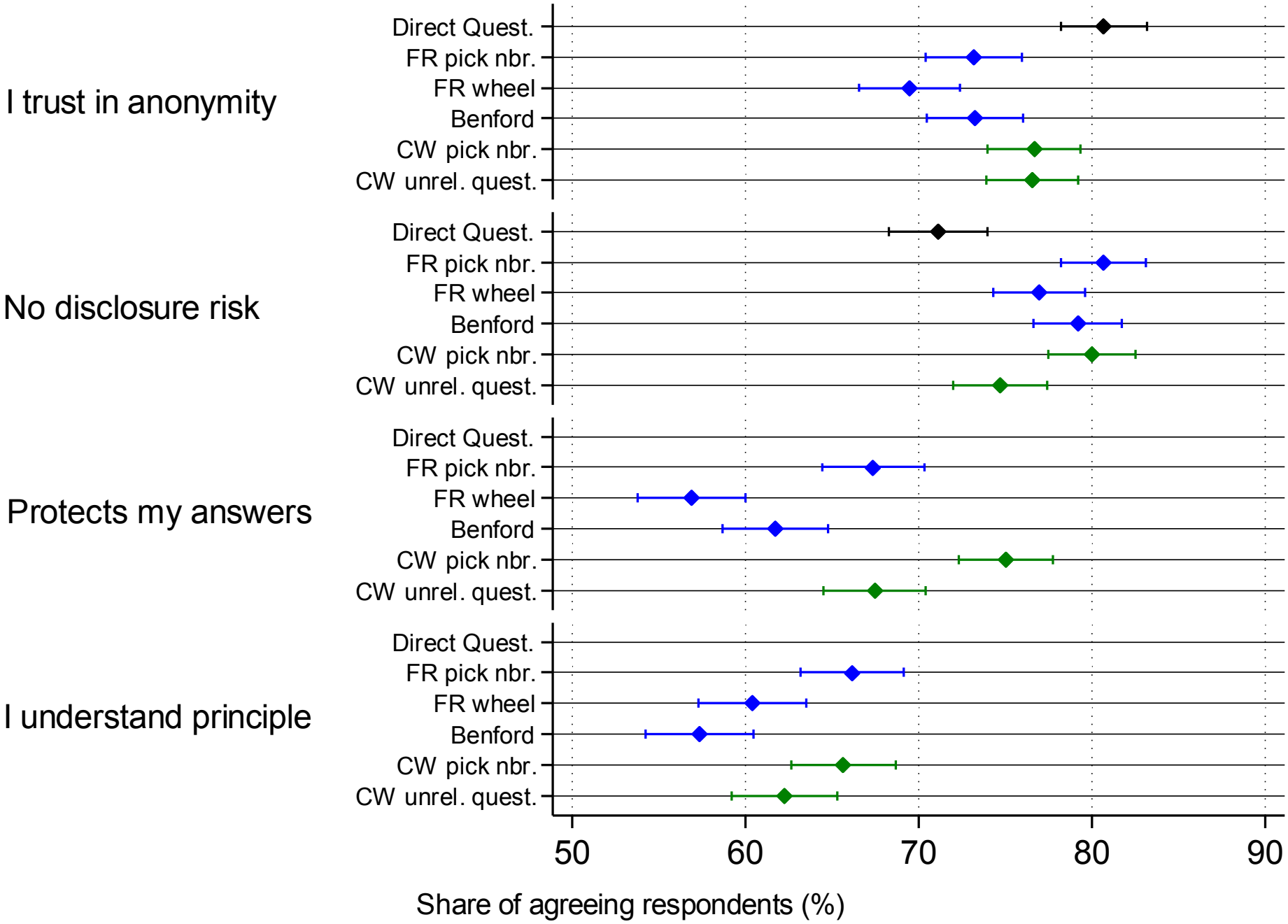
Experimental conditions

- DQ: direct questioning
  >
- RRT wheel: forced response RRT using virtual random wheel
  >
- RRT pick: forced response RRT using ‘pick a number’ method
  >
- RRT Benford: RRT using Benford distribution and innocuous questions
  >
  >>
- CW unr. ques.: Crosswise Model using unrelated questions

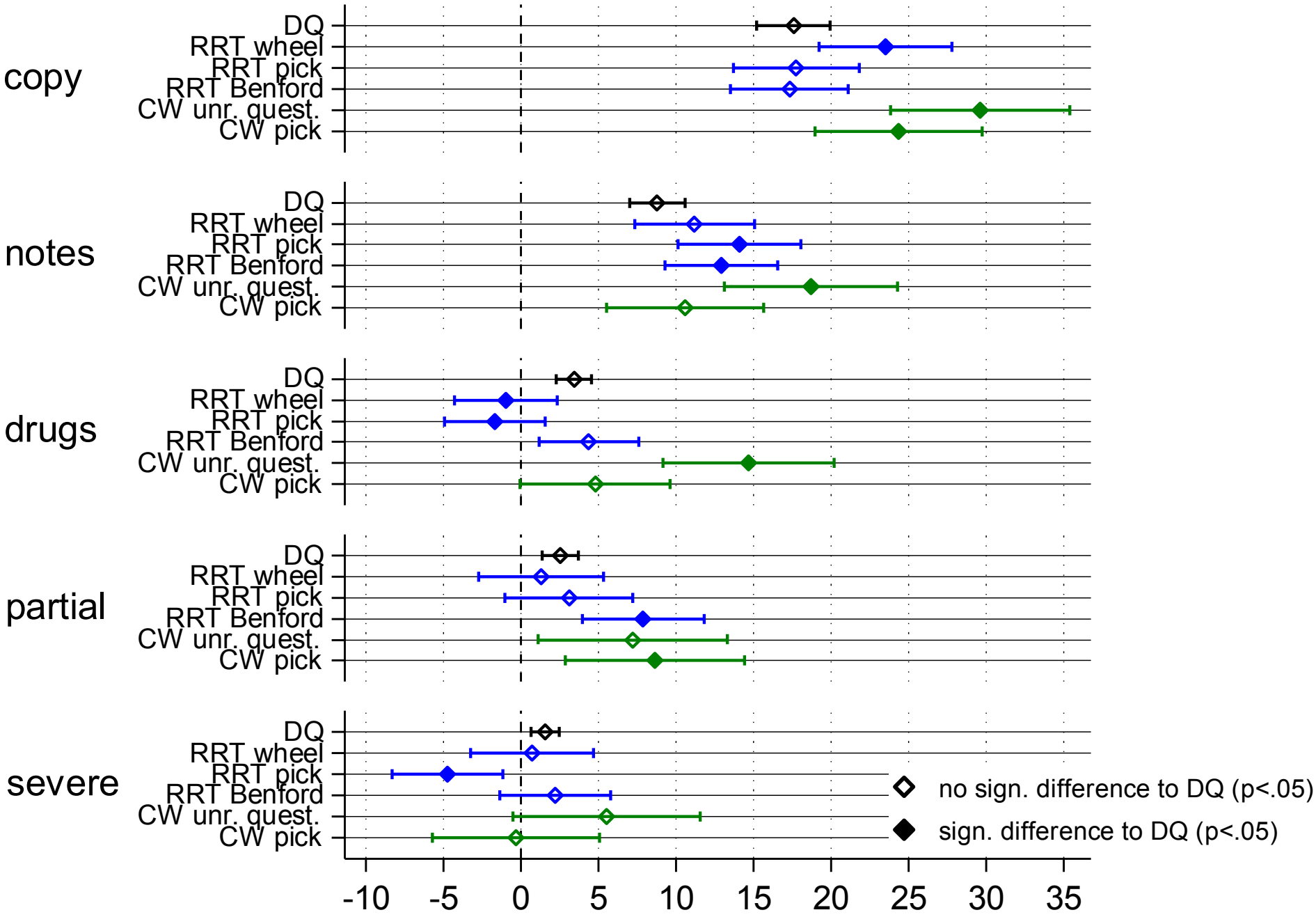
>

- CW pick: Crosswise Model using ‘pick a number’ method
  >

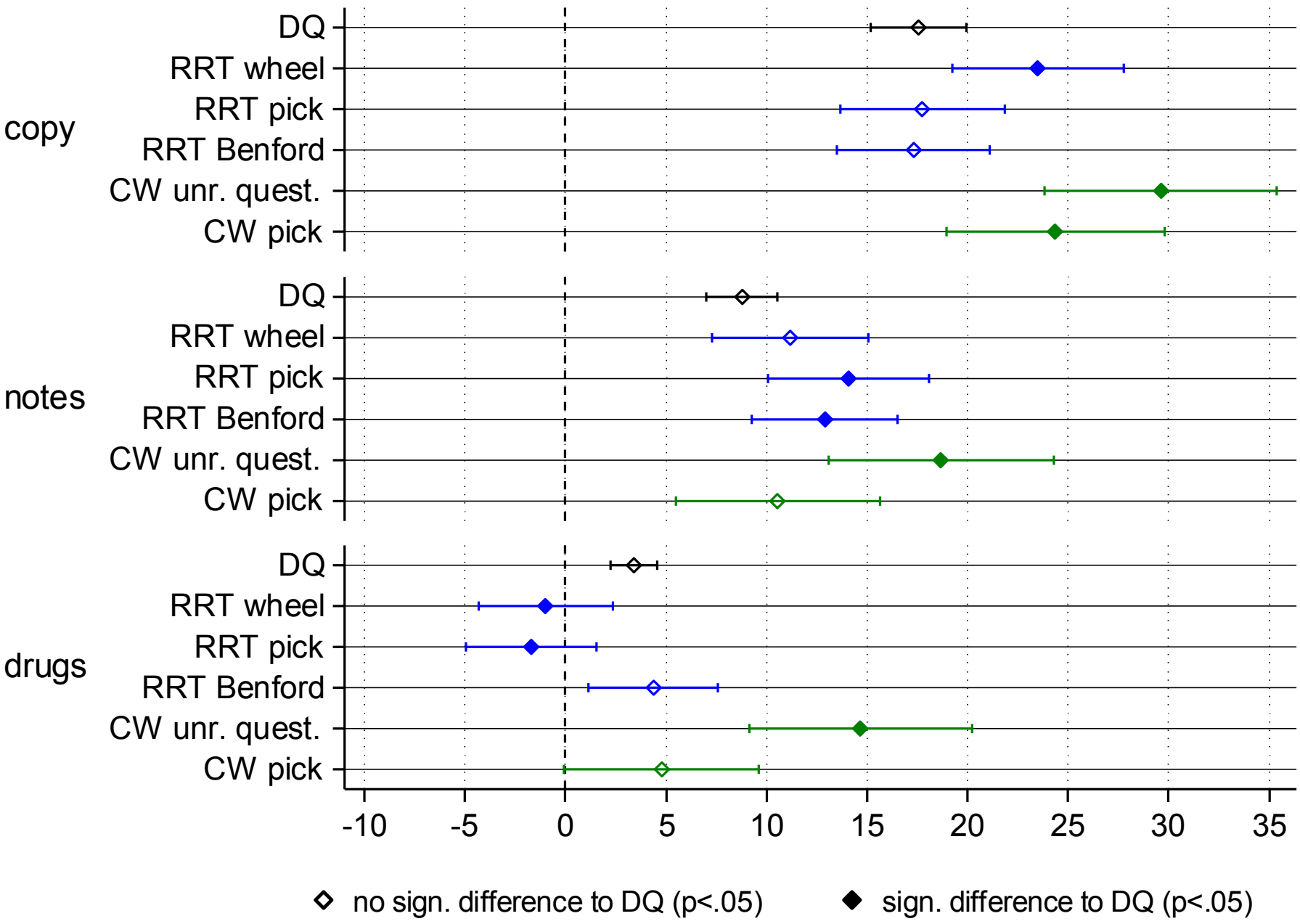
Respondents experience by experimental condition



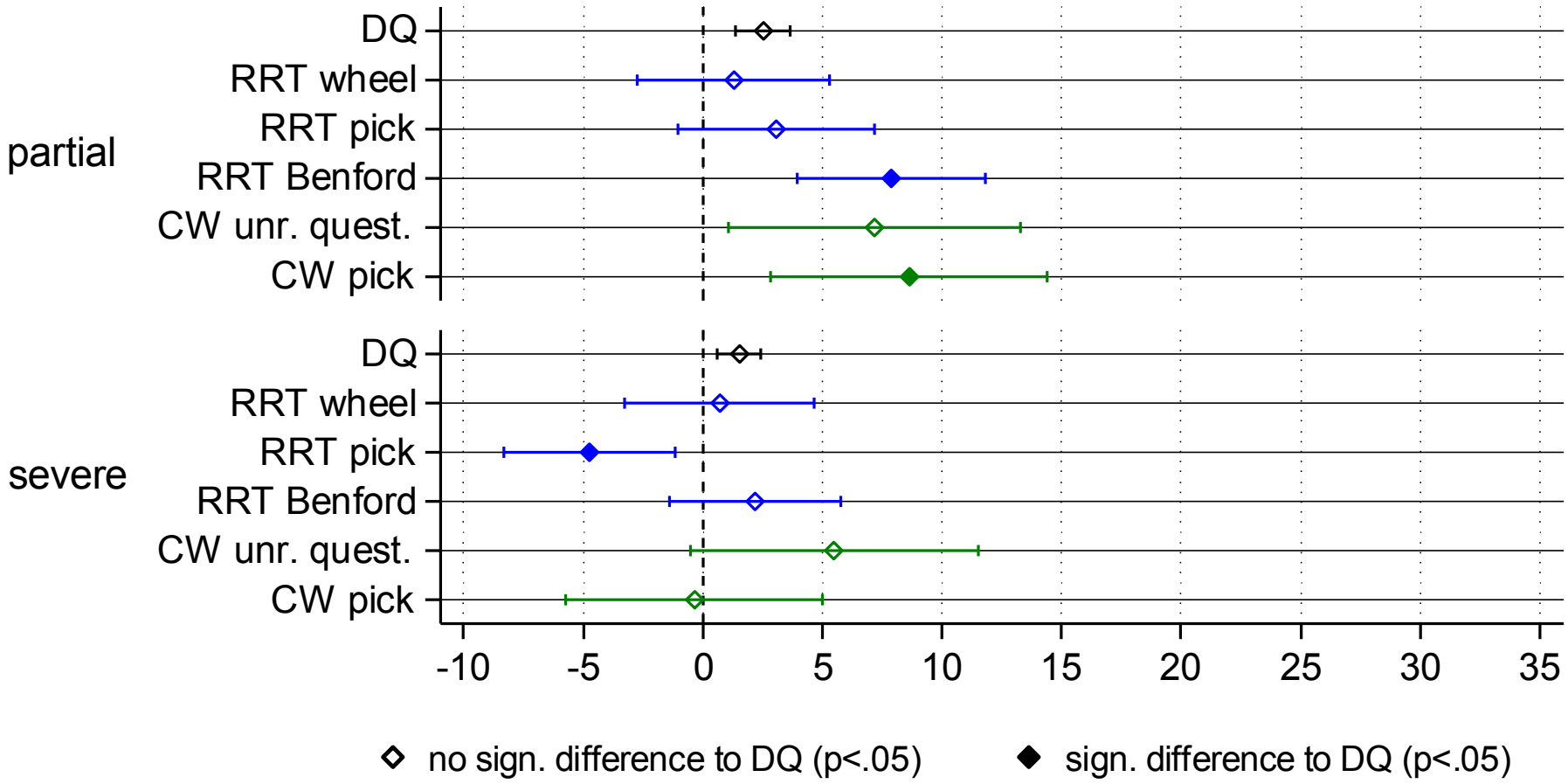
Point estimates and 95%-CI by experimental condition



Point estimates and 95%-CI by experimental condition



Point estimates and 95%-CI by experimental condition

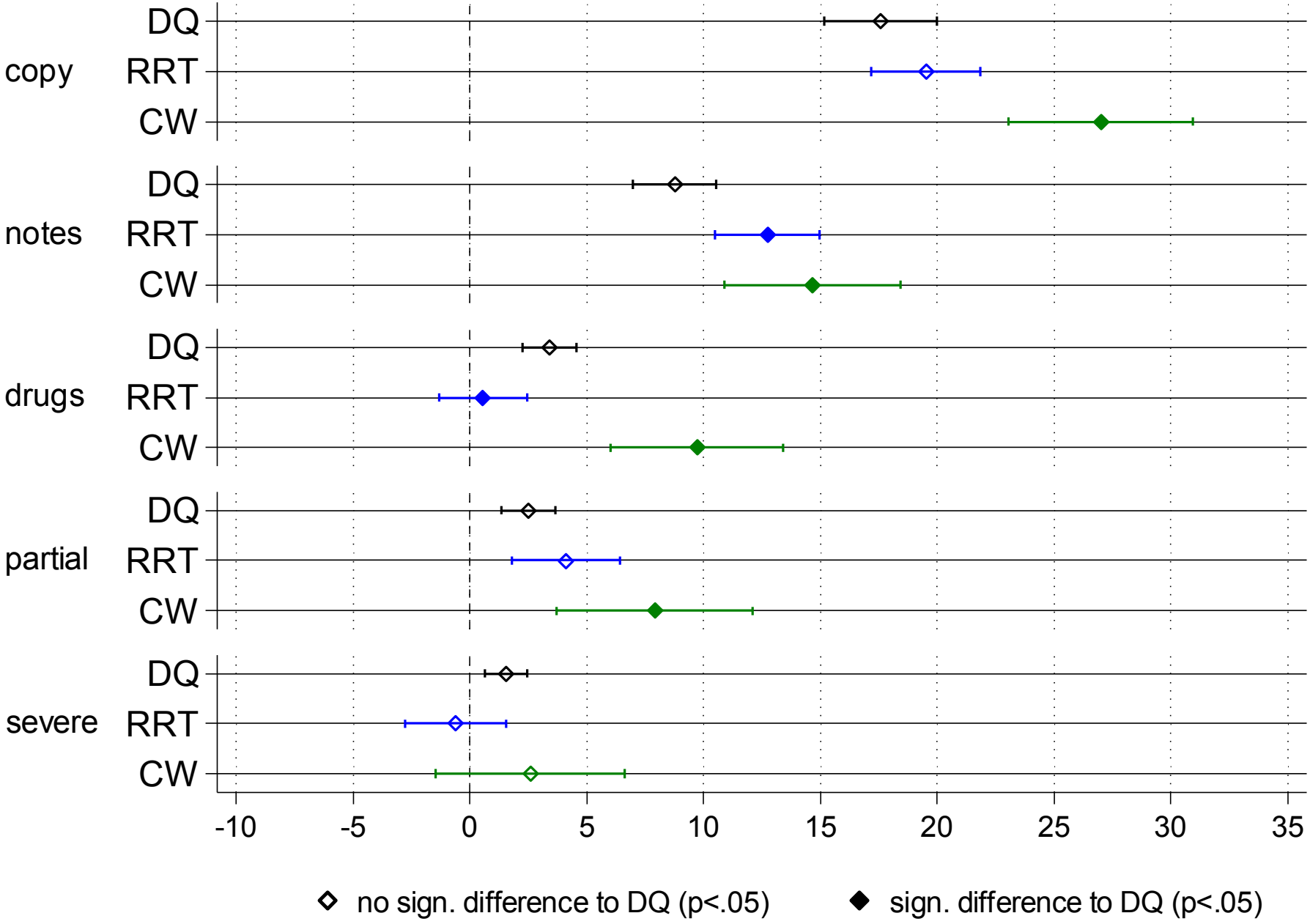


Prevalence estimates by experimental condition

	copy	notes	drugs	partial	severe
Direct questioning	17.6	8.8	3.4	2.5	1.5
	(1.2)	(0.9)	(0.6)	(0.6)	(0.5)
RRT random wheel	23.5	11.2	-1.0	1.3	0.7
	(2.2)	(2.0)	(1.7)	(2.0)	(2.0)
RRT pick a number	17.8	14.1	-1.7	3.1	-4.8
	(2.1)	(2.0)	(1.7)	(2.1)	(1.8)
RRT Benford	17.3	12.9	4.4	7.9	2.2
	(1.9)	(1.8)	(1.6)	(2.0)	(1.8)
CM unrelated question	29.6	18.7	14.7	7.2	5.5
	(2.9)	(2.8)	(2.8)	(3.1)	(3.1)
CM pick a number	24.4	10.6	4.8	8.6	-0.4
	(2.8)	(2.6)	(2.5)	(2.9)	(2.7)
Observations	5734	5735	5719	4232	4230

Standard errors in parentheses

Prevalence estimates aggregated



Prevalence estimates aggregated

	copy	notes	drugs	partial	severe
Level					
DQ	17.6	8.8	3.4	2.5	1.5
	(1.2)	(0.9)	(0.6)	(0.6)	(0.5)
RRT	19.5	12.7	0.6	4.1	-0.6
	(1.2)	(1.1)	(1.0)	(1.2)	(1.1)
CM	27.0	14.6	9.7	7.9	2.6
	(2.0)	(1.9)	(1.9)	(2.1)	(2.1)
Difference					
RRT - DQ	1.9	3.9**	-2.9*	1.6	-2.1
	(1.7)	(1.4)	(1.1)	(1.3)	(1.2)
CM - DQ	9.4***	5.9**	6.3**	5.4*	1.1
	(2.4)	(2.1)	(2.0)	(2.2)	(2.1)
N	5734	5735	5719	4232	4230

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Determinants of sensitive behavior

Randomized response logistic regression

	copy	notes	drugs	partial	severe
ETH (ref. UniBE)	-0.155	0.192	-0.542*	0.324	1.047
	(0.104)	(0.142)	(0.265)	(0.307)	(0.714)
Semester (log)	0.074	0.272	-0.046	0.026	-0.611
	(0.164)	(0.213)	(0.360)	(0.342)	(0.537)
Nbr. exams/papers (log)	0.601***	0.306*	-0.289	0.359	0.229
	(0.112)	(0.136)	(0.213)	(0.202)	(0.404)
Perceived risk	-0.015***	-0.023***		-0.010	-0.009
	(0.004)	(0.006)		(0.006)	(0.013)
Risk attitude	0.070**	0.094**	0.186**	0.175*	-0.020
	(0.025)	(0.033)	(0.066)	(0.083)	(0.134)
Procrastination	0.196***	0.211**	0.172	0.271	-0.031
	(0.050)	(0.067)	(0.128)	(0.153)	(0.323)
Stress at exams	0.106*	0.076	0.451***		
	(0.050)	(0.066)	(0.135)		
DQ	ref.	ref.	ref.	ref.	ref.
RRT	0.127	0.421**	-0.579	0.652	-12.812
	(0.117)	(0.152)	(0.445)	(0.340)	(668.330)
CM	0.499***	0.546**	1.105***	1.197**	1.126*
	(0.139)	(0.192)	(0.277)	(0.364)	(0.526)
Constant	-4.637***	-5.152***	-5.611***	-6.268***	-3.699*
	(0.418)	(0.602)	(0.926)	(1.124)	(1.557)
N	5663	5662	5673	4189	4186

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Summary

- The Crosswise Model clearly outperforms DQ (if we accept the ‘more-is-better’-assumption)
  - An exception is the last item (severe plagiarism) with a very low prevalence.
- RRT, on the other hand, does not yield higher estimates than direct questioning.
  - One reason might be the ‘self-protective no’ bias, which prevents respondents to say ‘yes’ if advised to do so by the randomizing device.

Methodological conclusions

- RRT does not seem to be a good method for self-administered surveys. Although we put a lot of effort into pretesting and finding good implementations, no convincing evidence could be found that RRT yields more valid estimates than DQ. (With RRT 'Benford' performing somewhat better than the other RRT implementations.)
- CM is a promising alternative, since it does not suffer from some of the deficiencies of RRT ("self-protective no" bias, complexity).
- Improvement of RRT estimates is possible by correcting for cheating respondents not complying with instructions (not shown). Such estimates, however, have low efficiency

Substantive conclusions

(based on combined results from CM)

- A substantial proportion of students have cheated on an exam (copying: about 25 percent, crib notes: about 15 percent)
- Using drugs to enhance exam performance is not uncommon (10 percent)
- Rates for partial plagiarism (using a passage from someone else's work without providing proper citation) are 8 percent. The prevalence of severe plagiarism (hand in someone else's work) is 3 percent.
- These numbers may not seem too high, but we have to keep in mind that they most likely still underestimate the real prevalence.

Thank you!

References

Coutts, E. und B. Jann. 2011. Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT), *Sociological Methods & Research*, Vol. 40, 196-193.

Coutts, Elisabeth, Benn Jann, Ivar Krumpal, and Anatol-Fiete Näher. 2011. "Plagiarism in Student Papers: Prevalence Estimates Using Special Techniques for Sensitive Questions." *Journal of Economics and Statistics 231:749-760.*

Edgell, S. E., S. Himmelfarb und K. L. Duchan. 1982. Validity of forced responses in a randomized-response model, *Sociological Methods & Research*, Vol. 11, 89-100.

Fox, J. A. und P. E. Tracy. 1986. *Randomized response: A method for sensitive surveys*, Newbury Park, CA: Sage.

Holbrook, A. L. und J. A. Krosnick. 2010. Measuring Voter Turnout By Using The Randomized Response Technique: Evidence Calling Into Question The Method's Validity, *Public Opin Q*, Vol. 74, 328-343.

Jann, B., J. Jerke und I. Krumpal. 2011. Asking Sensitive Questions Using the Crosswise Model: Some Experimental Results, *Public Opinion Quarterly*.

Lensvelt-Mulders, G. J. L. M. und H. R. Boeije. 2007. Evaluating compliance with a computer assisted randomized response technique: a qualitative study into the origins of lying and cheating, *Computers in Human Behavior*, Vol. 23, 591-608.

Locander, W., S. Sudman und N. Bradburn. 1976. An Investigation of Interview Method, Threat and Response Distortion, *Journal of the American Statistical Association*, Vol. 71, 269-275.

Peeters, C. F.W. 2006. Measuring politically sensitive behavior. Using probability theory in the form of randomized response to estimate prevalence and incidence of misbehavior in the public sphere: a test on integrity violations. Amsterdam: Faculty of Social Sciences, Vrije Universiteit Amsterdam.

Snijders, C., and J. Weesie. 2008. "The online use of randomized response measurement." Paper presented at General Online Research 2008, Hamburg, Germany.

van der Heijden, P. G. M., G. van Gils, J. Bouts und J. J. Hox. 2000. A Comparison of Randomized Response, Computer-Assisted Self-Interview, and Face-to-Face Direct Questioning. Eliciting Sensitive Information in the Context of Welfare and Unemployment Benefit, *Sociological Methods and Research*, Vol. 28, 505–537.

Warner, S. L. 1965. Randomized-response: A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, Vol. 60, 63-69.

Wolter, F., and P. Preisendörfer. 2011. "Asking Sensitive Questions: An Evaluation of the Randomized Response Technique vs. Direct Questioning Using Individual Validation Data." unpublished.

Yu, J.-W., G.-L. Tian, M.-L. Tang (2008). Two new models for survey sampling with sensitive characteristic: design and analysis. Metrika 67(3): 251-263.

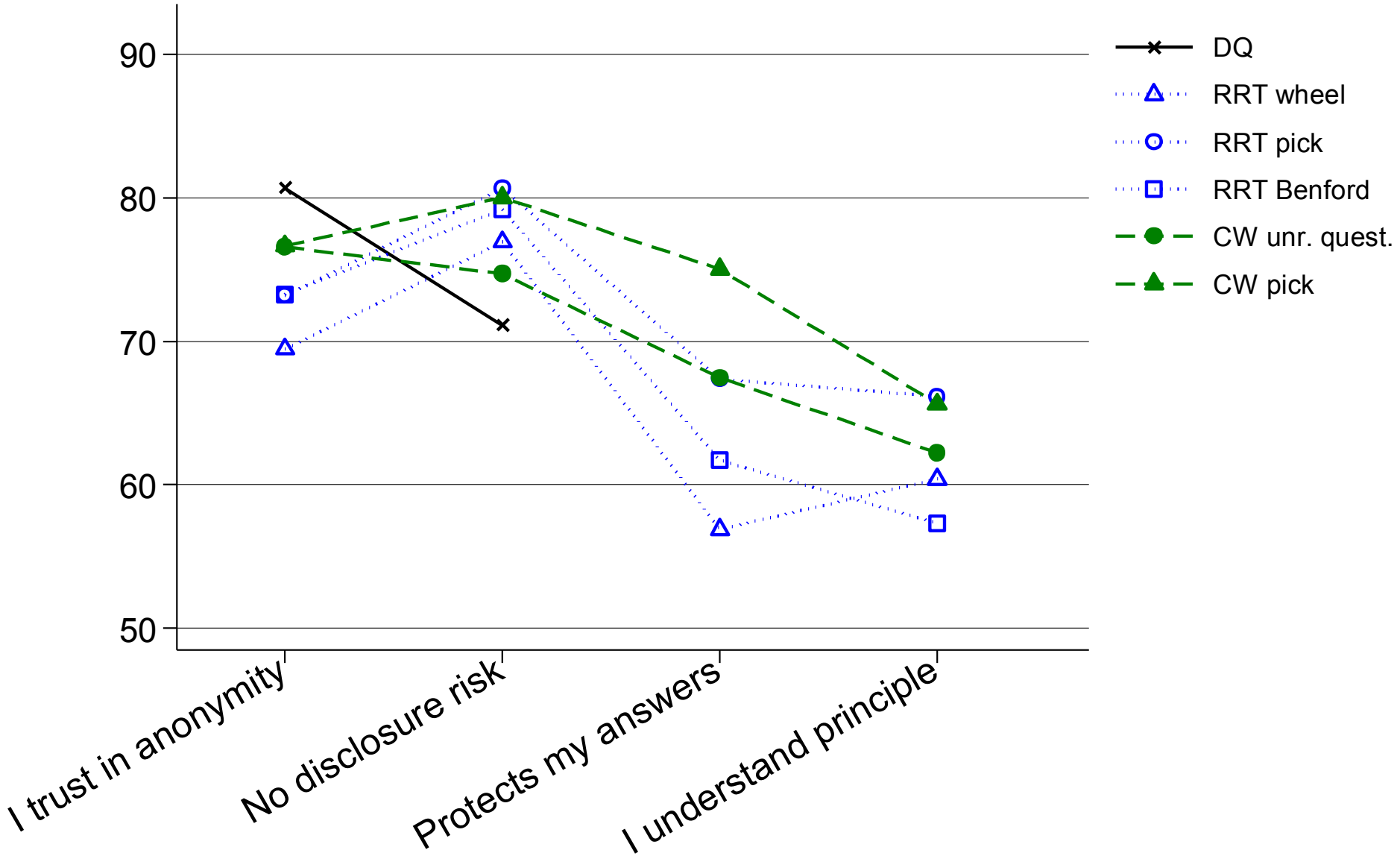
Appendix: Items

Question wording

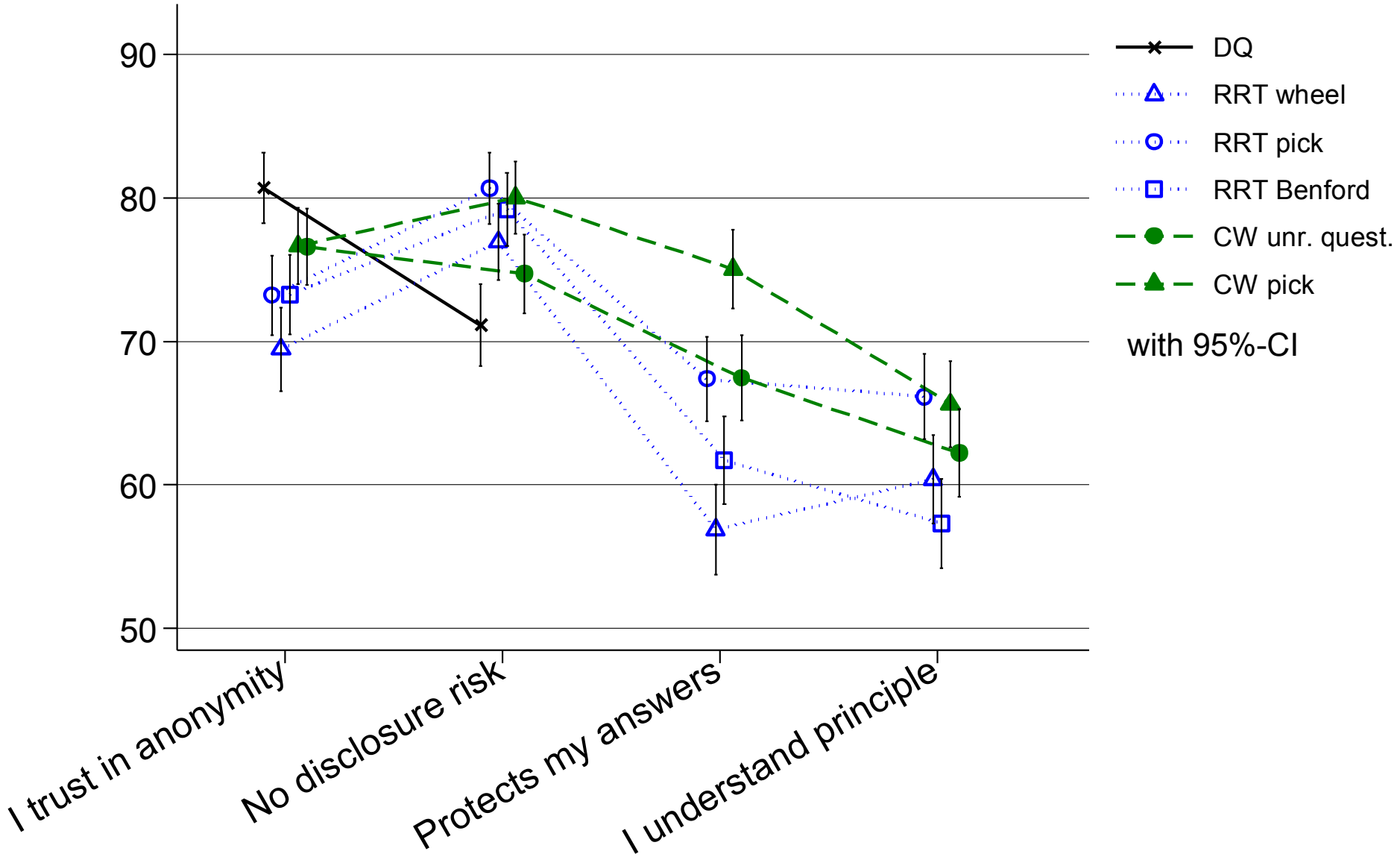
Item	Frageformulierung
Abschreiben	Haben Sie während Ihrem Studium jemals während einer Prüfung von Mitstudierenden abgeschrieben?
Spicken	Haben Sie während Ihrem Studium jemals unerlaubterweise einen Spickzettel (auch Handy-, Taschenrechner-Notizen und Ähnliches) in einer Prüfung verwendet?
leistungsfördernde Substanzen („Doping“)	Haben Sie während Ihrem Studium jemals rezeptpflichtige Substanzen/Medikamente eingenommen, um Ihre Leistung an Prüfungen zu steigern?
Teilplagiat	Haben Sie während Ihrem Studium jemals bei einer eingereichten Arbeit bewusst eine ganze Textpassage aus einem fremden Werk übernommen, ohne diese als Zitat zu kennzeichnen?
Vollplagiat	Haben Sie während Ihrem Studium jemals einen Grossteil einer Arbeit durch eine andere Person schreiben lassen oder eine fremde Arbeit als Ihre eigene ausgegeben?

Appendix: additional graphs & tables

Respondents experience by experimental condition



Respondents experience by experimental condition

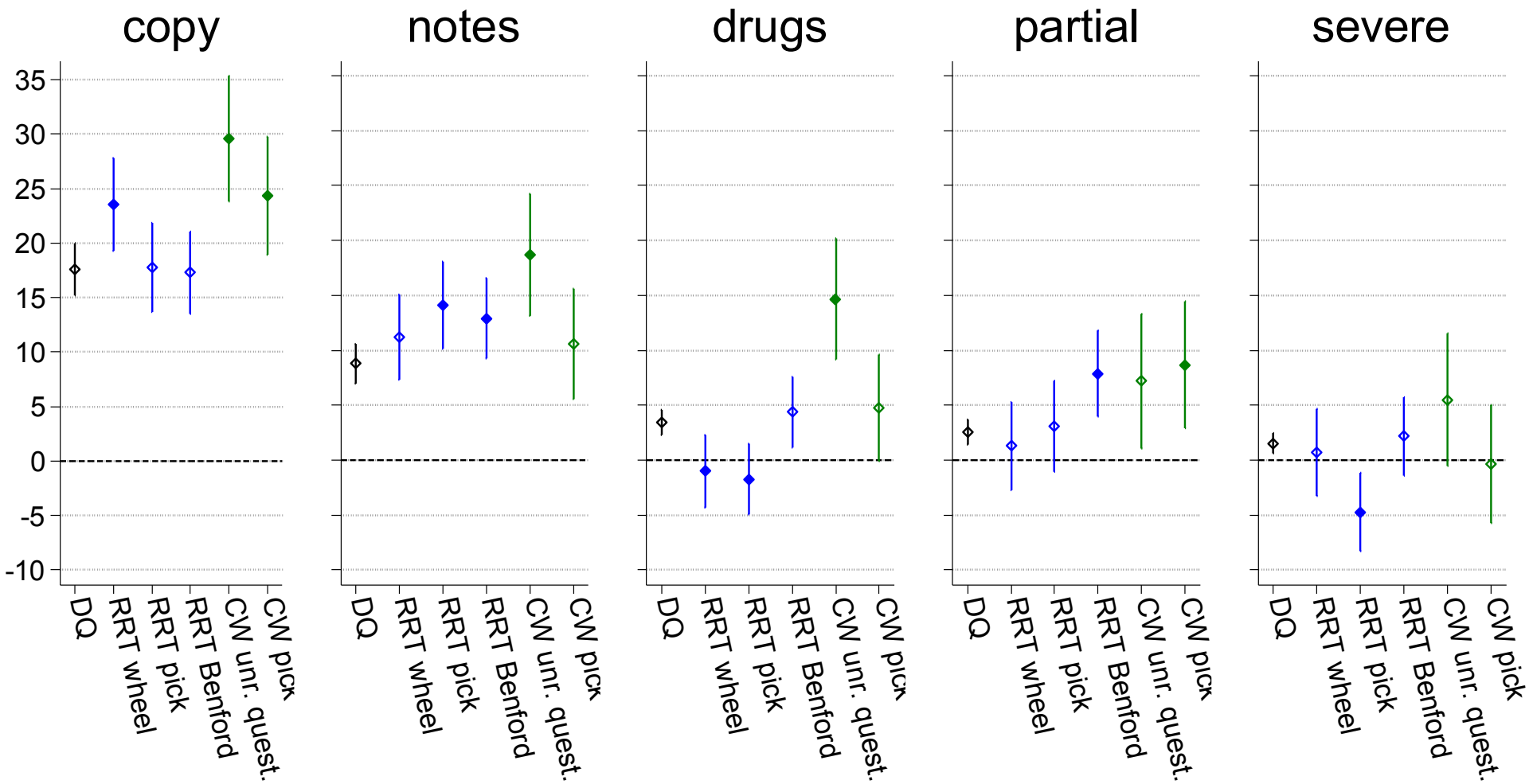


Prevalence estimates and differences to DQ

	copy	notes	drugs	partial	severe
Level					
Direct questioning	17.6	8.8	3.4	2.5	1.5
	(1.2)	(0.9)	(0.6)	(0.6)	(0.5)
RRT Benford	17.3	12.9	4.4	7.9	2.2
	(1.9)	(1.8)	(1.6)	(2.0)	(1.8)
RRT random wheel	23.5	11.2	-1.0	1.3	0.7
	(2.2)	(2.0)	(1.7)	(2.0)	(2.0)
RRT pick a number	17.8	14.1	-1.7	3.1	-4.8
	(2.1)	(2.0)	(1.7)	(2.1)	(1.8)
CM unrelated quest~n	29.6	18.7	14.7	7.2	5.5
	(2.9)	(2.8)	(2.8)	(3.1)	(3.1)
CM pick a number	24.4	10.6	4.8	8.6	-0.4
	(2.8)	(2.6)	(2.5)	(2.9)	(2.7)
Difference					
RRT rand. wheel - DQ	5.9*	2.4	-4.4*	-1.2	-0.8
	(2.5)	(2.2)	(1.8)	(2.1)	(2.1)
RRT pick number - DQ	0.2	5.3*	-5.1**	0.6	-6.3***
	(2.4)	(2.2)	(1.8)	(2.2)	(1.9)
RRT Benford - DQ	-0.3	4.1*	1.0	5.4*	0.7
	(2.3)	(2.1)	(1.7)	(2.1)	(1.9)
CM unr. quest. - DQ	12.0***	9.9***	11.2***	4.7	4.0
	(3.2)	(3.0)	(2.9)	(3.2)	(3.1)
CM pick number - DQ	6.8*	1.8	1.4	6.1*	-1.9
	(3.0)	(2.7)	(2.5)	(3.0)	(2.8)
Observations	5734	5735	5719	4232	4230

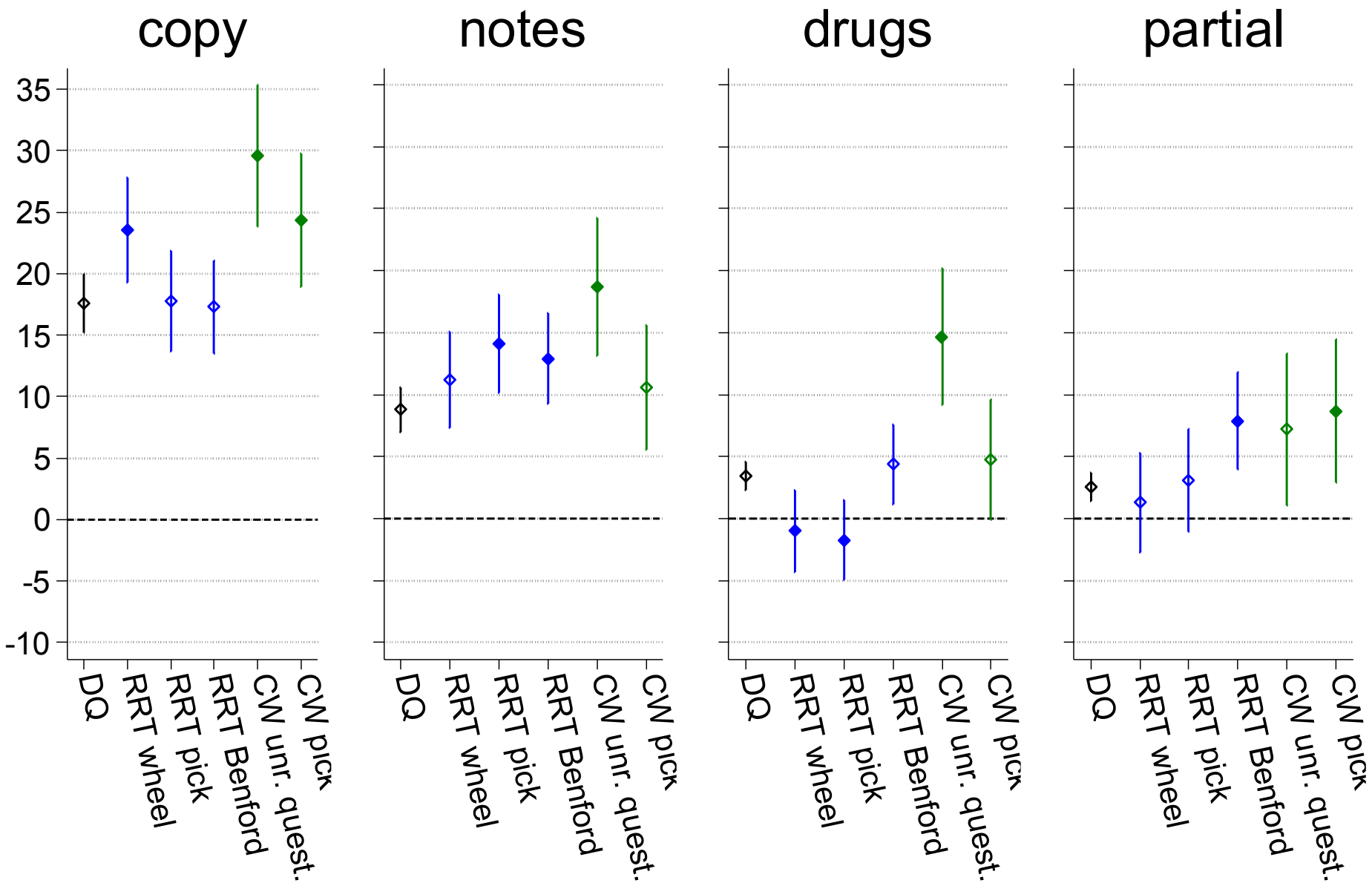
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Point estimates and 95%-CI by experimental condition



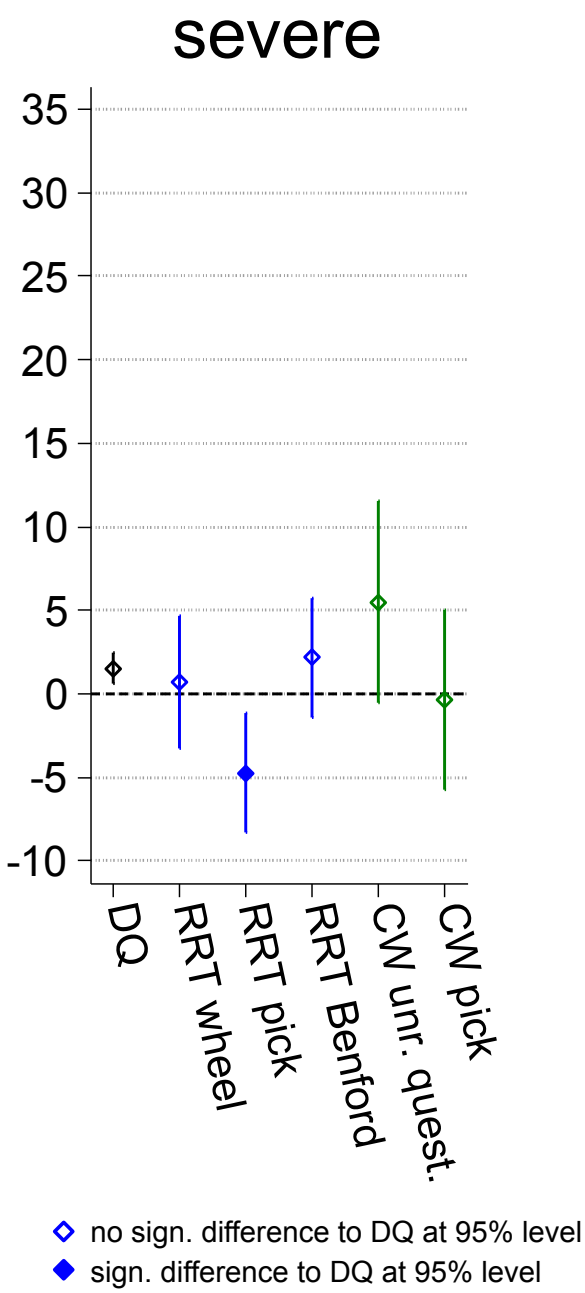
◇ no sign. difference to DQ at 95% level
◆ sign. difference to DQ at 95% level

Point estimates and 95%-CI by experimental condition

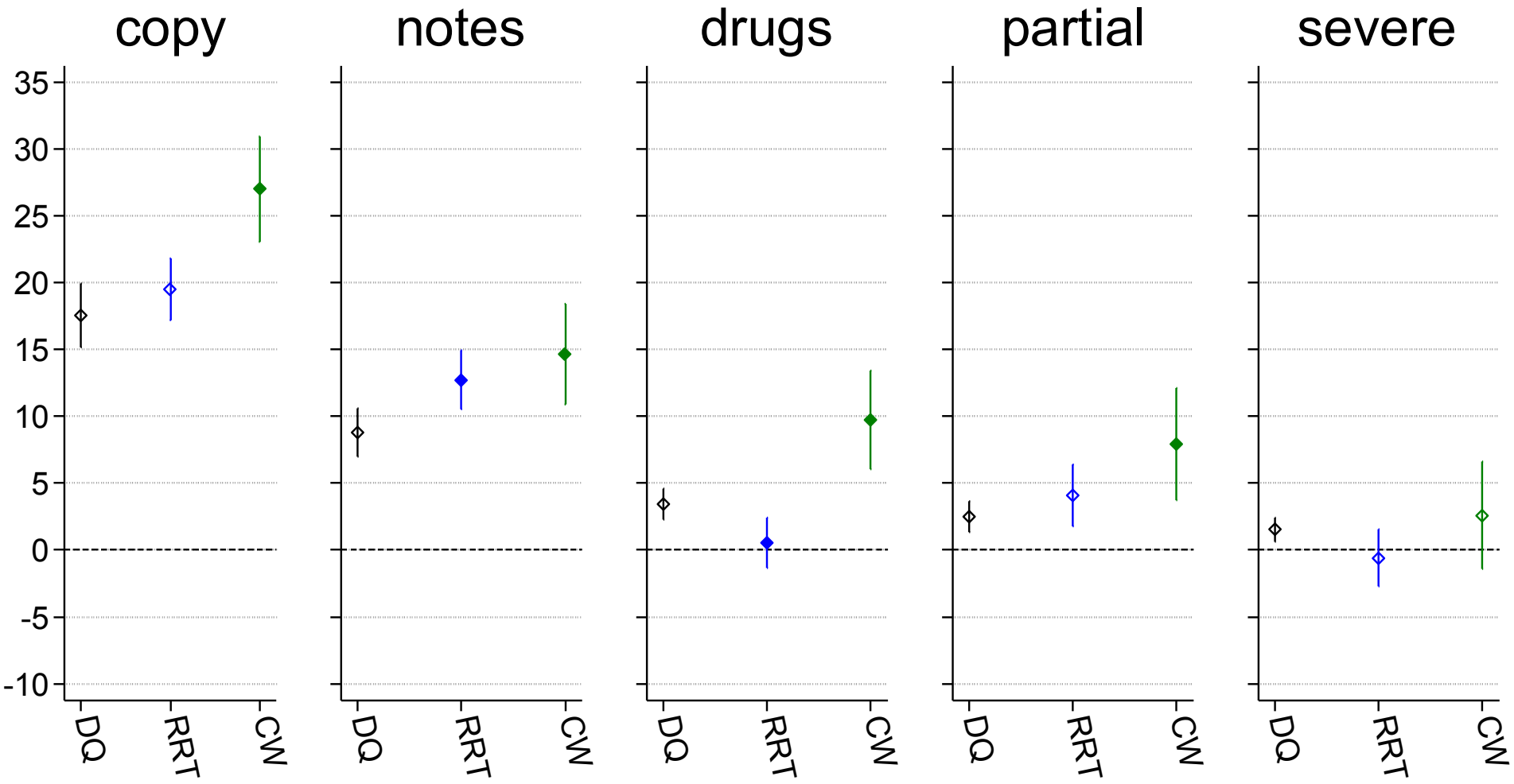


◇ no sign. difference to DQ at 95% level
◆ sign. difference to DQ at 95% level

Point estimates and 95%-CI by experimental condition



Prevalence estimates aggregated



◇ no sign. difference to DQ at 95% level
◆ sign. difference to DQ at 95% level