Towards a Game-theoretic Notion of Incoercibility

joint work with Dominique Unruh

CISPA Center for IT-Security, Privacy and Accountability

SAARLAND UNIVERSITY
COMPUTER SCIENCE

Disclaimer

- towards, as in: not there yet.

- started as M.Sc. thesis, extended for Crypto 2011

- found flaw afterwards

- (absolute) definition is incomplete, fix is unsound

- new proposal for relative notion, but how to tell it is correct?

Motivation: voting schemes

- you know this introduction :)
- for me, personally: what are the things we expect from voting? Can we learn more about pen-and-paper elections?
- won't talk about universal/individual verifiability

From secrecy to coercion resistance

- **vote-privacy:** protocol does not reveal your vote

- **receipt-freeness:** voter cannot produce receipt

- **coercion-resistance:** voter cannot produce receipt, even if interacting with coercer

From secrecy to coercion resistance

Andrea Ypsilanti, chairman of social-democratic party in Hess, 2008

asked/pressured MPs to take photograph of ballot.

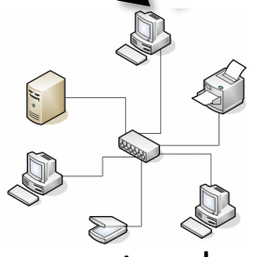
Motivation: game theory

- existing definitions (KTV 2010, UM-Q 2010): there is a *deception strategy* which
- achieves the good thing (e.g., votes for I)
- is indistinguishable from a voter which "obeyes"

UC/c

- aesthetically pleasing for people who like UC

$\forall$ 
adversary

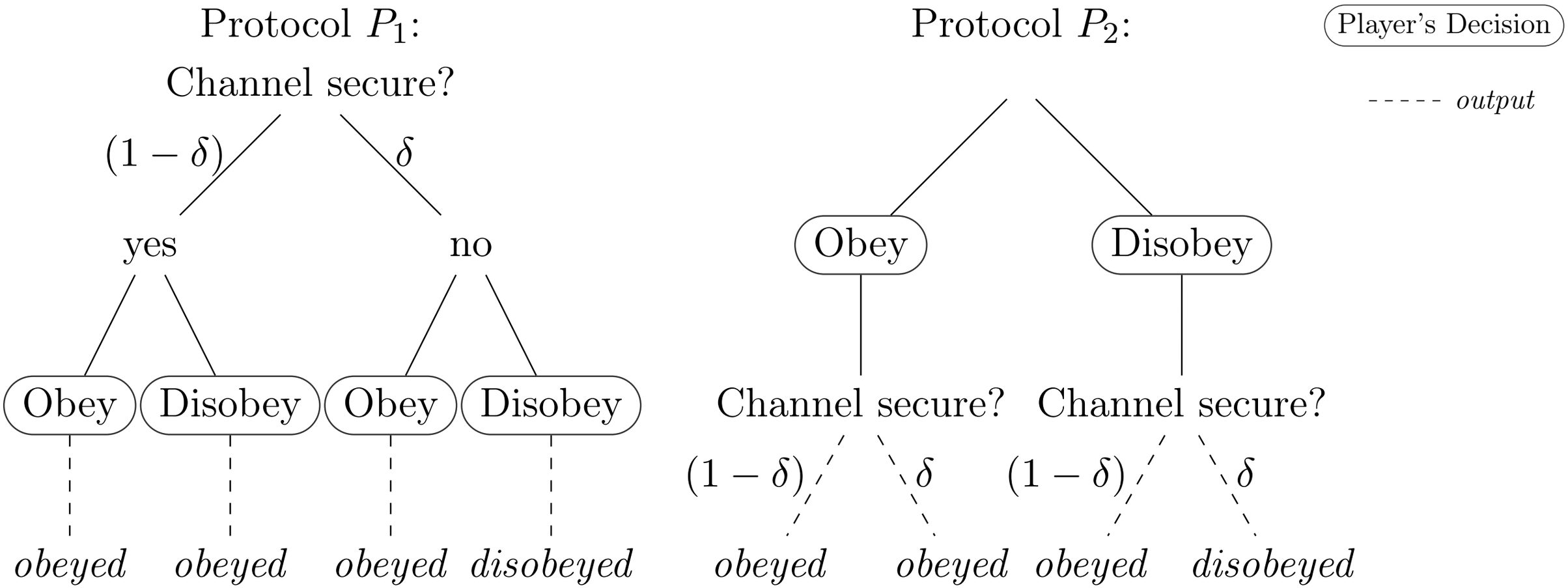

protocol

$\approx_\delta$

$\exists$ 
simulator

ideal machine

Announced randomness



Game-theoretic framework

- Two player player one-shot game with TMs as strategies:
  - players: P,C
  - output of game: outcome x punishment
  - utility functions for P/C, u: outcome → $\mathbb{R}$
  - punishment cost: p: punishment → $\mathbb{R}$

Equilibria

- most famous: Nash equilibrium
- best response to M_C: M_P,s.t. $u(M_C,M_P)\geq u(M_C,M_P')$
- everyone plays the best response to everyone else

- approach: coercion is in equilibrium
- best response to an effective coercer is obeying
- best response to an obeying player is: not coercing….

Best response for player

- deception strategy may depend on coercer:
  - best-response is a nice measure
  - but goal should still be reached
  - define disarmed/threat-utility as expected value of
    - u^{dis}: based on outcome
    - u^{thr}: u^{dis} - punishment

Harmless coercer

- a coercer is harmless iff best-response with and without threat have same disarmed utility

- Example: u(Clinton)=5 u(Trump)=0 u(Stein)=10

- protocol incoercible if for all (reasonable) utility functions, all coercers are harmless.

Allen Wertheimer

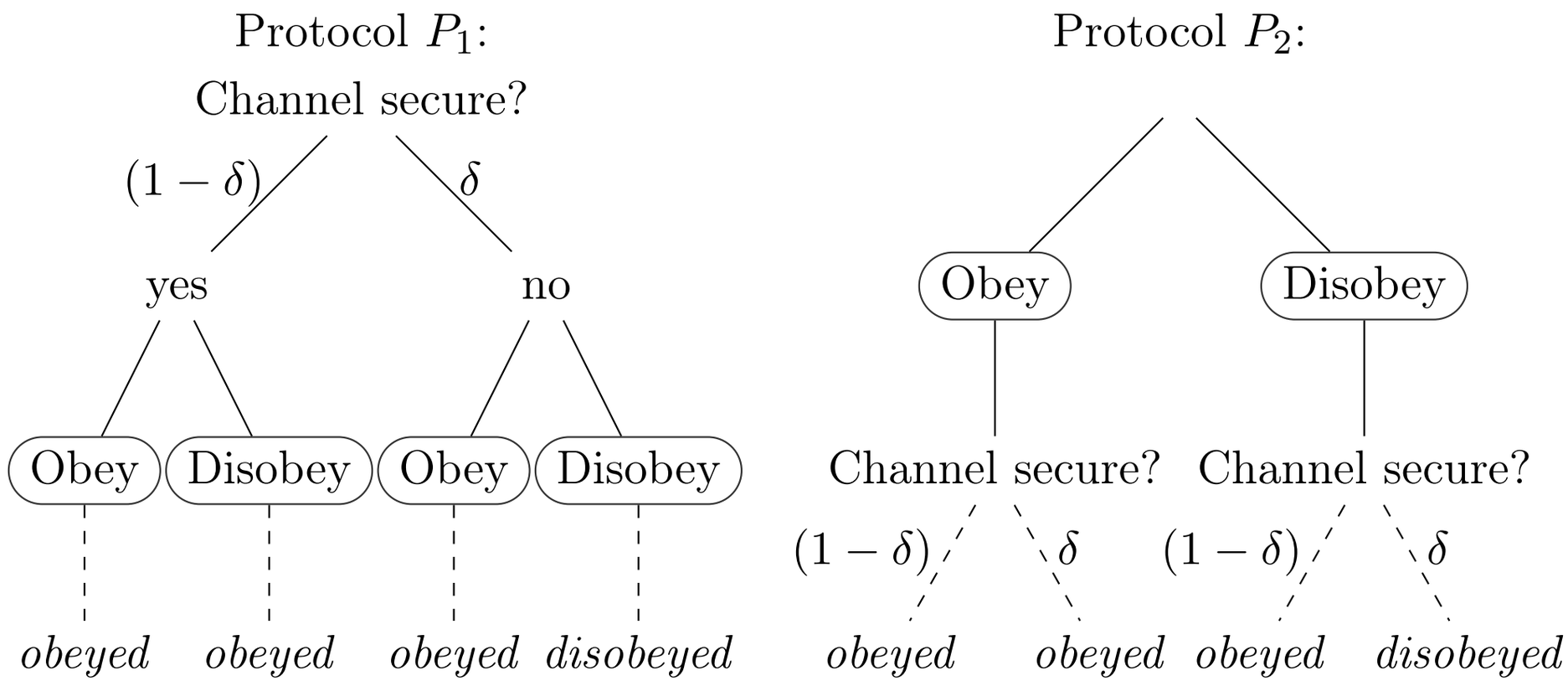
A. *The Proposal Prong.* A threatens B by wrongfully (viz. without moral justification) making B a proposal such that, unless B complies, B will be in a worse position than B was otherwise entitled to expect to be; and

B. *The Choice Prong.* B is morally justified in complying (e. g., since there is no reasonable alternative) and does comply with A's proposal.

$$\text{for all } M_P \in BR^{thr}(M_C) \quad M'_P \in BR^{dis}(M_C) \quad U_P^{dis}(M'_P, M_C) - U_P^{dis}(M_P, M_C) < \delta$$

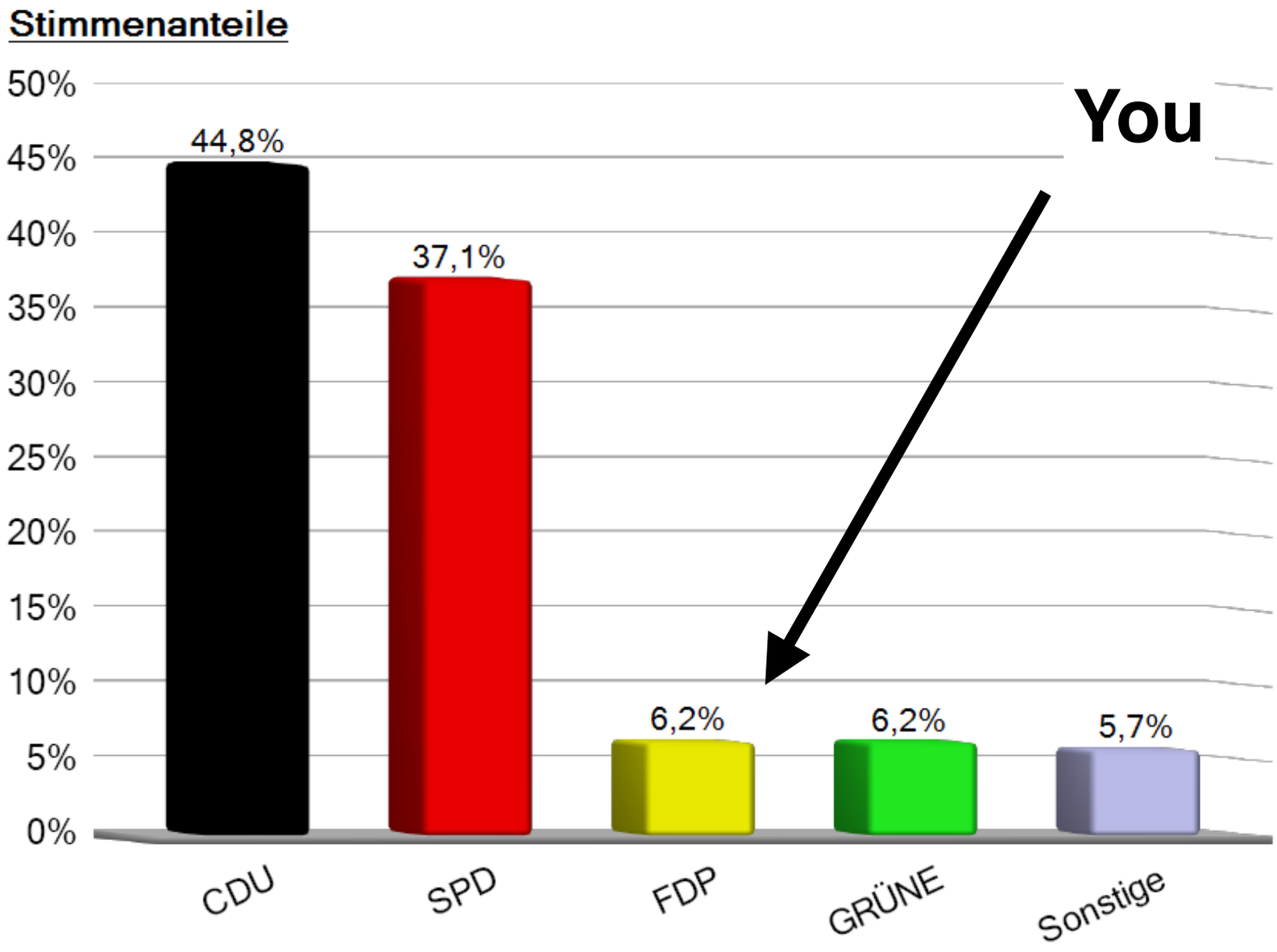
Results

- definition grounded in Wertheimer's Definition
- relative variant: P δ-as-good-as Q if for same utility function, same loss of disarmed utility (up to δ)
- connection to UC/c: if P emulates ideal voting functionality, then it is 0-as-good-as it.
- connection to KTV: if P is δ-KTV-incoercible, and u_{dis} smaller w, then min{w,δ}-incoercible (absolute).
- counter-example KTV: coercer makes his vote dependent via secret
- analysis of ideal voting in election district Saarbrücken: value of objective needs to be higher than 4% of punishment.



	G_1	G_2	Thm 10
$1 < w$	0	0	δ
$\delta < w \leq 1$	$\delta \cdot w$	0	δ
$w \leq \delta \leq 1$	$\delta \cdot w$	w	w

Analysis Saarbrücken



Problem: legal coercion

- coercer can vote, too.
- punishment by voting (incompleteness).
  - C votes Stein and punishes if P sends "Hello", otherwise Trump
  - threat-best-response sends Hello, disarmed-best-response not.
  - disarmed best-response has worse pay-off

Problem: legal coercion

$$\min_{M_P, M_C \text{ with } M_P \in BR^{dis}} U_P^{dis}(M_C, M_P)$$

$$\leq \min_{M_P, M_C \text{ with } M_P \in BR^{thr}} U_P^{dis}(M_C, M_P).$$

- legal coercion masks actual coercion: if voting for Trump is bad enough, receipt-giving protocol counts as incoercible (unsoundness..?).
- coercer can vote Trump to force P into providing receipt
- counter-example for indist.-based notions does not fly anymore

Proposal f. relative notion

- P is as-good-as Q, if
  - for all M_C and $M_P \in BR^{thr}(M_C)$ in P
    - there exist M_C' and $M_P' \in BR^{thr}(M_C')$ in Q
    - s.t. the outcome is (often enough) the same
- reflexive, transitive, but if Q is disarmed version of P, legal coercion still masks illegal coercion

Conclusion

- ?
- game-theoretic foundation for simulation-based notions seem achievable, framework has proven very useful.
- quantifying incoercibility in actual districts is interesting — large variations in size of voting districts in Germany could lead to problems
- approach: rational defender, irrational attacker seems ok (to me) can add economics, budget etc., too.
- maybe masking is ok? look only at utility functions where actual punishment is more severe than punishment by voting behaviour

Thank you for your attention.

Ask away!