# Where's Alice?: Applied Kid Crypto Meets Provable Security

Ryan Henry
University of Calgary
ryan.henry@ucalgary.ca

Alyssa Tory
University of Calgary
alyssa.tory@ucalgary.ca

Sophie Henry
Herons Crossing School

Isabella Henry
100 Worlds Daycare

Samantha Henry
Unaffiliated

## ABSTRACT

In this short paper, we revisit the celebrated Naor–Naor–Reingold (NNR) protocol for "[convincing] people you know where Waldo is without revealing information about his location". We observe that, despite oft-repeated claims to the contrary, the NNR protocol is neither *zero-knowledge* nor a *proof of knowledge*. We propose a slightly more elaborate version that is both of these things—but still eminently suitable for children's playdates (and the classroom).

## CCS CONCEPTS

• **Security and privacy** → **Privacy-preserving protocols**; *Cryptography*; • **Social and professional topics** → *Children*; *Adolescents*; *K-12 education*; *Adult education*.

## KEYWORDS

Zero-knowledge proofs; kid cryptography; bare-handed protocols; computer science education

## 1 INTRODUCTION

Zero-knowledge is at once a compelling and intimidating notion, promising—as if by magic—elegant solutions to countless "obviously unsolvable" computer security and privacy problems. This short paper arose out of the first author's experiences with teaching zero-knowledge to a general audience of undergraduate students (notably including the second author) using the celebrated *Naor–Naor–Reingold (NNR)* construction [8] for "[convincing] people you know where Waldo is without revealing information about his location" as an icebreaker.

Zero-knowledge protocols are becoming a crucial component of modern privacy-preserving systems, including cryptocurrencies, electronic voting schemes, anonymity systems, and more; however, they are also one of the hardest concepts to teach to laypersons because of their counterintuitive nature. If we expect wide adoption of privacy-preserving systems, we need a way to explain concepts like zero-knowledge to policymakers and the general public.

While Naor et al. never claimed that their construction is *zero-knowledge (ZK)* or a *proof of knowledge (PoK)*, a dogmatic belief that it is both ZK and a PoK has attained veritable "folklore status" among *kid crypto* enthusiasts. As we explain in the sequel, NNR is emphatically not ZK nor is it a PoK;[1] yet as a pedagogical device, we argue that it remains an instructive and gentle starting point for introducing students to the inscrutable world of ZK proofs and PoKs. Indeed, the very reasons NNR fails to formally qualify as ZK or a PoK ultimately serve to reinforce subtle aspects of the formal definitions. Moreover, it turns out that some relatively modest tweaks suffice to fortify NNR into a *bona fide* "bare-handed" Σ-protocol that is both special (general) zero-knowledge and special sound—all the while retaining its "playdate-friendly" allure.

This short paper serves three functions: (i) it describes and cryptanalyzes a provably secure NNR variant; (ii) it guides educators in how to incorporate NNR and its provably secure variant into introductory lessons on ZK proofs and PoKs; and (iii) it announces three "free culture"–licensed Waldo-esque puzzles for use in the classroom and beyond.

## 2 NAOR–NAOR–REINGOLD, REVISITED

NNR prescribes a single-move two-party protocol intended to be run face-to-face between a prover Peggy and a verifier Victor. The protocol is *bare-handed* in the same vein as *bare-handed voting protocols* [9], not requiring the use of computers or other sophisticated technology. The common input to Peggy and Victor is a *Where's Waldo?* puzzle;[2] Peggy seeks to convince Victor that she has found Waldo—without betraying Waldo's whereabouts.

Beyond the puzzle itself, Peggy and Victor require an opaque cover (e.g., a sheet of cardboard) trimmed to (approximately; see below) twice the dimensions of the puzzle and with a Waldo-sized *viewport* cut out of its center. Peggy places the cover over the puzzle, carefully aligning it so that Waldo (and, to the extent possible, only Waldo) is visible through the viewport.

On one hand, if Peggy has indeed found Waldo, then so aligning the puzzle is trivial (albeit tedious[3]). In other words, NNR exhibits *(perfect) completeness*: if Peggy is not lying, then Victor is always convinced.

On the other hand, if Peggy has *not* found Waldo, then she cannot properly align the cover; indeed, upon doing so, Peggy could place some sort of beacon within the viewport, remove

[1] Its inventors further disavow any notion that it is "particularly educational"; on this point, we respectfully disagree.

[2] *Where's Waldo?* is a series of children's puzzle books in which each two-page spread illustrates a busy scene featuring hundreds of characters engaged in a variety of amusing acts. Somewhere amid the crowded scene is Waldo—sporting his distinctive red-and-white striped shirt and bobbled hat—and you are asked to scour the detailed illustrations to find him. An example puzzle from the original book is provided in Figure 1.

[3] Naor et al. [8, §2.2] recommend that "to actually execute it [Peggy] should place her finger on Waldo while navigating the [opaque cover]".

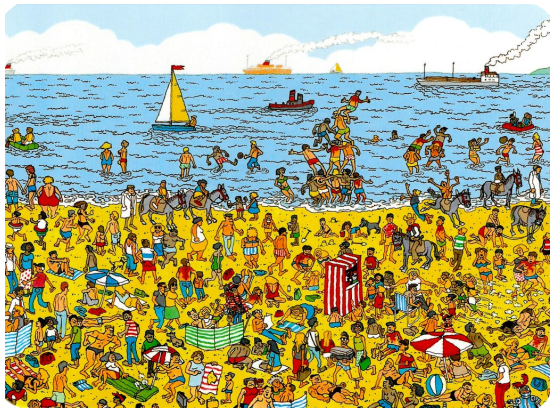


**Figure 1: "On the Beach", scene from Where's Waldo? [5].**

the cover, and then locate the beacon to reveal Waldo's whereabouts in the puzzle. Intuitively, the preceding contraposition seems to imply that NNR also exhibits *(perfect) soundness*: if Peggy is lying, then Victor is not convinced.[4]

The dimensions of the opaque cover are carefully selected so as to conceal the entire puzzle (apart from the image of Waldo in the viewport), thus preventing Victor from gaining perspective about Waldo's location relative to the puzzle's extents. More precisely, suppose that (i) the puzzle is rectangular with width $W$ and height $H$ and that (ii) Waldo is confined to an axis-aligned rectangle of width $w$ and height $h$. Then the puzzle extends beyond the viewport horizontally by at most $W-w$ and vertically by at most $H-h$ in any direction, irrespective of Waldo's whereabouts. Thus, an axis-aligned viewport of width $w$ and height $h$ centred within an opaque cover of width at least $2W-w$ and height at least $2H-h$ ensures that Victor at worst learns something about Waldo's immediate surroundings as viewed through the viewport. This suggests that NNR is (at least approximately) *zero-knowledge*; i.e., *Victor learns (essentially) nothing beyond that Peggy is not lying.*

## 2.1 Zero-knowledge proofs of knowledge

The analysis of NNR in the preceding section well illustrates the intuition behind claims that NNR is ZK and a PoK. To make such claims rigorous, however, it is necessary to evaluate them against the formal definitions of these notions (as presented in Definitions 2.1 and 2.2, respectively).

*"Zero-knowledge" proofs.* The notion of "zero-knowledge" proofs was formalized in a seminal 1985 paper by Goldwasser, Micali, and Rackoff [4], who observed that *interaction* (i.e., the ability of Peggy and Victor to exchange messages back and forth) and *probabilism* (i.e., the ability of Peggy and Victor to toss coins to decide which messages to send) together enable Peggy to establish facts via statistical arguments that yield no information beyond the veracity of her claims. The property of "yielding no information" is formulated in the *simulation paradigm* [3] by showing that Victor can efficiently "simulate"—without even speaking to Peggy—all the information he might obtain throughout the interaction.

*Definition 2.1 (Zero-knowledge; semi-formal).* A two-party protocol between Peggy and Victor is *zero-knowledge (ZK)* if there is an efficient procedure with which, given only the common input, Victor can sample "counterfeit" interaction transcripts from *the same distribution* as "genuine" transcripts describing his real interactions with Peggy.

We model Victor's efficient counterfeiting procedure as a probabilistic polynomial-time (PPT) algorithm (the "*simulator*") and refer to the act of sampling counterfeit transcripts as *simulating* interactions with Peggy.

Recall that NNR has only a single (deterministic) move: Peggy presents Victor with a fully formed "proof" and Victor merely confirms that Waldo's likeness appears in the viewport. More commonly, ZK protocols consist of several consecutive moves, with messages from Victor attempting to expose any deception on the part of Peggy.[5] In contrast with the original NNR construction, our provably secure variant is a so-called $\Sigma$-*protocol* [2] comprising three moves: (i) Peggy announces a randomized "commitment"; (ii) Victor tosses a challenge coin; and (iii) Peggy responds to Victor's challenge. Victor accepts the proof as valid if and only if Peggy's response satisfies some verification procedure involving Peggy's randomized commitment and Victor's random challenge from the opening two moves.

*Proofs "of knowledge".* ZK captures the idea that Victor learns nothing beyond the veracity of Peggy's claims. Often, what Peggy claims is first-hand "knowledge" of some *witness* to a (purported) fact. Precisely what it means for Peggy to "know" such a witness may at first blush appear nebulous, yet an influential 1992 paper by Bellare and Goldreich [1] provides a compelling characterization: Peggy "knows" $x$ if and only if she can (be modified to) efficiently *compute* it.

*Definition 2.2 (Proof of knowledge; semi-formal).* A two-party protocol between Peggy and Victor is a *proof of knowledge (PoK)* of $x$ if there is an efficient procedure that, by interacting with (and possibly "rewinding") Peggy, outputs $x$ with about the same probability as Peggy is able to convince Victor of her claims.

We model the efficient procedure from Definition 2.2 as an expected PPT algorithm (the "*knowledge extractor*") and refer to the act of using it to compute $x$ as *extracting* $x$ from Peggy. The knowledge extractor is allowed to "rewind" Peggy by instructing her to return to an earlier state, thereby seeing how she would have responded—*in the same move of the*

[4] Jumping ahead, we stress that this consequence need not hold *as the premise may be false*: a cheating Peggy might present Victor with a *simulation*!

[5] As it happens, the lack of random challenges from Victor in NNR is indicative of its security flaws. Anecdotally, we have observed that, after exposing students to interactive (provably ZK) protocols and then revisiting NNR, many students can intuit that the lack of messages originating from Victor is a red flag bearing further scrutiny.


**Figure 2: Image visible through the NNR viewport for 24 Where's Waldo? puzzles from Where's Waldo? The Phenomenal Postcard Book [6].**

*same conversation*—to several distinct challenges from Victor.[6] Apart from rewinding, the knowledge extractor interacts with Peggy as a "black box" through the same "interactive proof" interface she exposes when interacting with Victor. Thus, if interactions between Peggy and Victor are ZK, then rewinding is *necessary* to tease out the information needed to extract $x$ from Peggy.

## 2.2 Security analysis

In light of Definitions 2.1 and 2.2, we can now formally reassess claims that NNR is ZK and a PoK. As per Definition 2.1, whether NNR is ZK depends on Victor's (in)ability to *convincingly simulate* his interactions with Peggy. In this context, simulation entails producing counterfeit "transcripts" with Waldo's likeness visible through the viewport of the opaque cover. Assuming the viewport is rectangular, as suggested both by Naor et al. [8] and earlier description of NNR, the viewport in "genuine" transcripts necessarily displays elements of the scene that Victor cannot anticipate without having already (i) run the protocol with Peggy or (ii) located Waldo on the puzzle. Otherwise, Victor is resigned to sampling counterfeit transcripts from a distribution that is *perceptually distinguishable* from that of genuine transcripts, with the implication that **NNR is not ZK in the sense of Definition 2.1**. Figure 2 shows the image visible through the viewport when NNR is run over 24 different *Where's Waldo?* puzzles from *Where's Waldo? The Phenomenal Postcard Book* [6]. The abundance of variation is striking: not one pair of viewports from different puzzles in that book look alike.

[6] In an actual playdate, Peggy should refuse to rewind to an earlier state; i.e., knowledge extraction is best regarded as a *thought experiment*. This distinction between how Peggy *could* behave in theory versus how she *should* behave in practice is an example of what Middendorf and Shopkow refer to as a "learning bottleneck" [7], a near-ubiquitous conceptual stumbling point for beginning students. In our experience, having students pair up to manually run ZK protocols and their associated knowledge extractors helps them to conceptualize the role of knowledge extraction in more abstract settings. Looking forward, our NNR variant and associated puzzles are eminently suitable for this sort of classroom activity.

As per Definition 2.2, whether NNR is a PoK depends on the existence of a knowledge extractor that can compute Waldo's location by interacting with *and rewinding* Peggy. Because Peggy receives no messages from Victor during the interaction, rewinding endows the extractor with no capabilities beyond those also available to Victor. In practice, the additional context one obtains from Waldo's immediate surroundings in the viewport can greatly simplify the process of locating Waldo; however, modulo what is leaked by the failure of NNR to be ZK, a hypothetical knowledge extractor gains no benefit from interacting with Peggy, with the implication that **NNR is not a PoK in the sense of Definition 2.2**.

# 3 NAOR–NAOR–REINGOLD, RELOADED

Our NNR variant introduces *interaction* and *probabilism* while maintaining the same clever idea that underlies the original NNR construction. Yet even with the addition of these two ingredients, proving that the resulting construction is indeed ZK is possible only if we further assume that the puzzle given as common input satisfies two "invariants"; namely, that Waldo (or, at least, Waldo's head)

(1) has an a priori known shape, size, and orientation; and
(2) is unobstructed by elements of the surrounding scene.

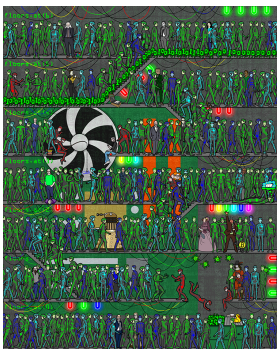
Taken together, the two invariants imply the existence of an opaque cover whose viewport, when properly aligned, exactly shows Waldo (or Waldo's head) *and not a pixel more*, thus depriving Victor of any information about the scene unfolding in Waldo's immediate vicinity. Unfortunately (as evident in Figure 2), published *Where's Waldo?* puzzles frequently violate both invariants. For reasons that will become apparent in the analysis, Peggy and Victor additionally require *four identical copies* of the puzzle, arranged in a 2-by-2 grid where each quadrant is a copy.[7]

The setup is similar to—but somewhat more elaborate than—that of the original NNR construction. Suppose that (i) the puzzle is rectangular with width $W$ and height $H$ and that (ii) Waldo (or Waldo's head) is confined to an axis-aligned rectangle of width $w$ and height $h$. Beyond the puzzle given as common input, Peggy and Victor require *three* opaque covers with different dimensions and different sized and shaped cutouts:

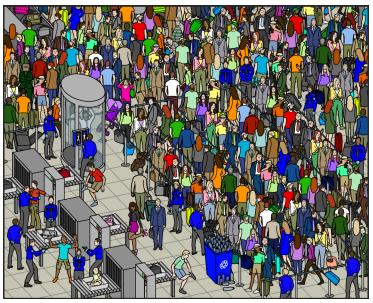
(1) the first cover is trimmed to *thrice* the puzzle dimensions with an (axis-aligned) rectangular viewport of width $W + w$ and height $H + h$ cut out of its center;
(2) the second cover is trimmed to *four times* the puzzle dimensions with a Waldo-shaped, -sized, and -oriented viewport cut out of its center (taking advantage of the invariants); and
(3) the third cover is a rectangle trimmed to a width of $W + w$ and a height of $H + h$ with no viewports cut out.

The interaction between Peggy and Victor is as follows.

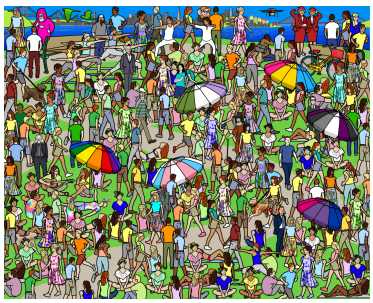
[7] Fortunately, the three Waldo-esque puzzles that we present in the last section each satisfy both invariants and are easily replicated into such a grid.



(a) The Matrix Maze



(b) Airport Security



(c) Alice in Friscoland

**Figure 3: Our three new Where's Alice? posters.**

**Peggy's announcement:** Peggy randomly aligns the first cover on top of the 2-by-2 puzzle grid so that the origin is located uniformly within the axis-aligned, puzzle-sized rectangle centered in the first cover. Next, she locates Waldo within the viewport and aligns the second cover so that Waldo (and only Waldo) is visible through the viewport. Finally, she aligns the third cover with the puzzle-sized viewport of the first cover. The resulting stack of covers and a puzzle is her randomized commitment.

**Victor's challenge:** Victor flips an unbiased coin to generate a challenge, either "heads" or "tails".

**Peggy's response:** Peggy responds to Victor's challenge as follows:

- if the challenge is "heads", then she lifts the third cover (revealing an image of Waldo); otherwise
- if the challenge is "tails", then she lifts the second and third covers simultaneously (revealing the poster's position relative to the first cover).

**Victor's verdict:** Victor accepts if either (i) the challenge was "heads" and Waldo is visible through the second cover's viewport or (ii) the challenge was "tails" and the poster is visible through the first cover's viewport.

Some remarks about this protocol are in order. The protocol relies on a 2-by-2 grid of puzzles to mimic the effects of modular reduction; in uniformly placing the first cover, Peggy is in effect uniformly shifting the poster cyclically, modulo its extents. The viewport on the first cover is cut to a width $W + w$ and height $H + h$ to account for shifts that happen to split Waldo into pieces. Specifically, enlarging the viewport by a Waldo-sized amount beyond the dimensions of the puzzle ensures that at least one "contiguous" copy of Waldo appears within the viewport, regardless of the random shift and Waldo's whereabouts. In particular, the first cover commits Peggy to a random 2-dimensional cyclic shift of the puzzle, while the second cover serves the same purpose as the cover in NNR; the third cover hides Waldo's location within the shifted puzzle. Victor eventually learns *either* shift amounts *or* Waldo's location within the shifted puzzle, but never both.

Notice also that the protocol has soundness error $\frac{1}{2}$. If the challenge is "heads", Peggy can fool Victor by aligning an image of Waldo (not from the puzzle) with the viewport of the second cover; however, this strategy will backfire if the challenge is "tails" since Victor will spot the image of Waldo on top of the shifted puzzle. Likewise, if the challenge is "tails", then Peggy can fool Victor by placing the second cover arbitrarily; however, this strategy will backfire if the challenge is "heads" since Victor will not see Waldo in the viewport. This means that Peggy and Victor must repeat the protocol several times to reduce the soundness error to acceptable levels.

The construction is ZK and a PoK; we omit proof of that claim here so that it may be assigned as homework.

## 4 "WHERE'S ALICE?" POSTERS

We created three custom security- and privacy-themed Waldo-esque puzzles, which we preview in Figure 3. Each puzzle satisfies the two invariants from Section 3; in fact, the posters feature an entire cast of characters that satisfy the invariants, including both fictional characters like Alice and Bob and prominent figures from the security and privacy research communities and computer science more broadly. The original Adobe Photoshop (`.psd`) sources are reconfigurable, allowing to relocate the cast of findable characters to produce new puzzles from a given scene. The posters are distributed freely under a Creative Commons Attribution 4.0 International (CC BY 4.0) license[8]. High-resolution copies (including the originals) are accessible via https://pr.iva.cy/alice.

*Acknowledgements.* We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) through grant RGPIN 2019-04821.

## REFERENCES

[1] Mihir Bellare and Oded Goldreich. On defining proofs of knowledge. In *Advances in Cryptology: Proceedings of the 12th Annual International Cryptology Conference (CRYPTO 1992)*, volume 740 of *Lecture Notes in Computer Science*, pages 390–420, Santa Barbara, CA, USA (August 1992).

[2] Ivan Damgård. On Σ-protocols. Lecture notes for CPT 2011, University of Aarhus BRICS, Aarhus, Denmark, March 2011.

[8] https://creativecommons.org/licenses/by/4.0/

[3] Oded Goldreich, Silvio Micali, and Avi Wigderson. How to play ANY mental game (Or a completeness theorem for protocols with honest majority). In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC 1987)*, pages 218–229, New York, NY, USA (May 1987).

[4] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof-systems (extended abstract). In *Proceedings of the 17th Annual ACM Symposium on Theory of Computing (STOC 1985)*, pages 291–304, Providence, RI, USA (May 1985).

[5] Martin Hanford. *Where's Waldo?* Candlewick Press, Somerville, MA, USA, 1987.

[6] Martin Hanford. *Where's Waldo? The Phenomenal Postcard Book*. Candlewick Press, Somerville, MA, USA, 2011.

[7] Joan Middendorf and Leah Shopkow. *Overcoming Student Learning Bottlenecks: Decode the Critical Thinking of your Discipline*. Stylus Publishing, LLC, Sterling, VA, USA, 2017.

[8] Moni Naor, Yael Naor, and Omer Reingold. Applied kid cryptography (Or how to convince your children you are not cheating). *Journal of Craptology*, 0(1):3, August 1998.

[9] Ben Riva and Amnon Ta-Shma. Bare-handed electronic voting with pre-processing. In *Proceedings of the 2007 USENIX/ACCURATE Electronic Voting Technology Workshop (EVT 2007)*, Boston, MA, USA (August 2007).