# Position Paper: A Theoretical Framework for General Cognition Evaluation of Cognitive Radios

Qi Dong†, Yu Chen†, Xiaohua Li†, Kai Zeng‡
†*Dept. of Electrical and Computing Engineering, Binghamton University, SUNY, Binghamton, NY 13902, USA*
‡*Volgenau School of Engineering, George Mason University, Fairfax, VA 22030, USA*
*E-mails: {qdong3, ychen, xli}@binghamton.edu, kzeng2@gmu.edu*

***Abstract*—Cognitive radio network (CRN) has been considered as the most promising solution for an efficient usage of wireless communication spectrum resource as the pervasive utility of Internet of Things (IoT) and other smart mobile devices. Such cognitive capabilities collectively define the intelligence of CRNs. While the capability of cognition and the intelligence are vital for CRNs, the quantitative study is largely an open area. Because of the structural complexity and the nature of multidisciplinary, there is not a theoretically reasonable method to evaluate the "effectiveness" and "intelligence" of cognitive radio models. In contrast, the human intelligence quotient (IQ) test provides a straightforward quantitative description of a person's capability in conducting certain types of tasks like logical thinking. Inspired by the well-known Cattell-Horn-Carroll (CHC) theory for human cognitive ability evaluation, in this paper a theoretical framework is proposed trying to apply the CHC model on cognitive radios. What this paper reports is the preliminary effort to lay down a theoretical framework for general cognition evaluation of cognitive radio or CRNs. By sharing the early stage conceptual idea with the colleagues in the smart cities community, we hope it will inspire more insightful discussions to accelerate the research in this important area.**

***Keywords*-Cognitive Radio Networks (CRNs), Model Evaluation, Intelligence Quotient (IQ), Cattell-Horn-Carroll (CHC) Theory.**

## I. Introduction

In decades of research towards efficient spectrum resource exploration and smart wireless system, "cognitive radio" is a smart solution to versatile radio environment and a revolutionary policy to distribute spectrum resource [16]. Due to the rigid spectrum allocation scheme regulated by governmental agencies, cognitive radio allows cognitive wireless communication devices extensively explore spectrum resource usage efficiency, without introducing interference to licensed users. The term "cognitive radio" is actually a broad topic that consists of a set of comprehensive technologies, including software-defined radio (SDR), wireless communication structure development, artificial intelligence (AI) and machine learning (ML). A well-designed cognitive radio network (CRN) aims to serve two purposes: to maximize the usage efficiency of spare spectrum resource as well as to protect the incumbent primary system from secondary network interference [9].

Because of the structural complexity and the nature of multi-disciplinary, there still lack of a theoretically reasonable structure to evaluate the "effectiveness" and "intelligence" of cognitive radio models. The difficulty originates from several aspects:

1) The wireless communication is conducted in a versatile open radio environment that varies too much to find a "proper" testing environment for cognitive radio;
2) The cognitive radio is functional in a cooperative way that usually works in a complex network environment, while the network can be constructed adopting many structures, such as hierarchical, distributed or heterogeneous formation;
3) The performance of cognitive radio differs from different numbers of cognitive devices and different position of each devices in a cluster;
4) It is still not clear what the most important performance metrics are out of so many choice of performance measurements; and
5) It is an open question how to define the "intelligence level" of cognitive radios.

From a traditional perspective of model evaluation method, it is easy to be overwhelmed by the great complexity of cognitive radio evaluation structure. For example, In a comprehensive cognitive evaluation model, the cognitive radio performance metrics are divided into three levels: node-level metrics, network-level metrics and application-level metrics [20]. In the node level, metrics are mostly focused on spectrum resource management and the general intelligence of cognitive radio nodes. In network level, metrics are generally focused on the overall performance and intelligence level of the whole network, such as network reliability, scalability, security and QoS. In application level, there are great challenges to define proper performance metrics for general applications.

In spite of the theoretical robustness of those traditional model based evaluation methods, it is not practical to design standard test batteries following that trace. First of all, too many metrics have been included in the model. Actually, some of them are highly correlated, which makes the model unnecessarily complicated. Secondly, some metrics' definitions are obscure and abstract, such as the classification of intelligence quotient (IQ) level for cognitive radio nodes as "infant", "toddler", "preschool", "child", "adolescent", "teenager", "adult". While some network metrics are not directly measurable such as "overall network IQ level", some others are not really meaningful without specifying network topologies. In addition, even if all the metrics are measurable and reasonable to cognitive radios, it is extremely difficult, if not impossible, to fabricate all possible radio environments, network structures and applications as test batteries to evaluate a cognitive radio technique.

An antidote to the dilemma of cognitive radio evaluation is to design standard benchmarks as test batteries, where the cognitive radio is treated as a black box. "Benchmarking the output as the intelligence level designation to cognitive radio" allows users to feed any types of cognitive radios into test batteries as the input [18]. A robust benchmarking theory should satisfy:

1) diversity to include "enough" testing scenarios;
2) completeness to measure critical performance metrics;
3) practically concise to maintain effective testing and framework design; and
4) output informative description of cognitive radio.

In the design of test batteries for benchmarking, a practical problem is the choice of design platforms. Basically, the platforms fall into two categories: software-based simulators and hardware-based emulators. Normally, the cognitive radio testing object can be a physical cognitive wireless device, or a non-specific cognitive radio technique proposal. Either "cognitive radio" forms shall be applicable to test batteries. Thus, the choice of two different classes design platforms is usually a search for a good trade-off. Software-based simulators, such as MATLAB, NS2, NS3, OmNET++ etc, provide flexibility and scalability on designing various testing scenarios, while hardware-based emulators, such as BEE2 [7], USRP [1], offer higher confidence and correctness of testing, but usually are not scalable. For example, a cognitive radio system evaluation methodology was proposed using Wireless Open-Access Research Platform (WARP) as the testing platform [18]. It is a promising solution but is very likely to be restricted by its limited scalability. The platform selection is beyond the scope of this paper, we encourage the interested readers consider the cognitive radio test battery as a combination deployment of both software-based simulator and hardware-based emulator.

The human IQ test provides a straightforward quantitative description of a person's capability in conducting certain types of tasks, *e.g.* logical thinking. Intuitively, it would also be helpful if the smartness of a cognitive radio can be quantitatively marked. Inspired by the well-known Cattell-Horn-Carroll (CHC) theory for human cognitive ability evaluation, in this paper a theoretical approach is proposed trying to apply the CHC model on cognitive radios. Following the proposed cognitive radio evaluation methodology, the designed test battery shall be beneficial to both researchers who want to design cognitive radio, and customers who want to purchase proper cognitive radio products.

This paper reports our preliminary effort to lay down a theoretical framework for a general cognition evaluation of cognitive radio or CRNs. By no means it is a completed work, instead, it is more appropriate to be considered as an early stage conceptual paper. By sharing the idea with colleagues in the smart cities community we hope it inspires more insightful discussions and encourages collaborations.

The rest of this paper is organized as follows. In section II, the background knowledge inspiring the proposed method is presented. Section III theoretically describes preparation of our evaluation model construction, and discuss some critical points when design test batteries. Section IV presents the theoretical design process. Section V concludes the paper with some discussions on potential future developments.

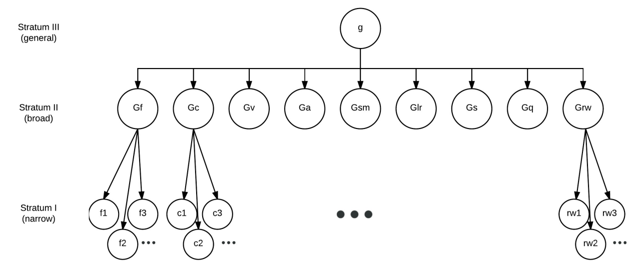


Figure 1: The CHC model of human cognitive taxonomy.

## II. Background Knowledge and Related Work

### A. CHC Model

Due to the high complexity of structure of human intelligence, there is not a universal agreed rule set of human behaviors. In process of searching ways to describe and measure human intelligence capabilities, the concept "psychometric taxonomic" was brought to shed some lights on human intelligence research. During the last few decades, the Cattell-Horn-Carroll (CHC) theory, a well-known human cognitive ability taxonomy theory, becomes a widely accepted comprehensive and empirically-based taxonomy of human cognitive ability [15]. It is the most comprehensive and empirically supported psychometric theory of the structure of cognitive abilities to data [10]. Based on the theory, a CHC model is constructed as a three-stratum hierarchical model to describe human intelligence as shown in Fig. 1. The model presents three levels of cognition: over 70 narrow abilities at stratum I, nine broad abilities at stratum II as shown in Table I, and a general ability $g$ at the apex of the hierarchical model as stratum III. This model divided human intelligence into several distinctive broad aspects, which can

Table I: The Stratum II of CHC model: broad ability terms.

| | |
|---|---|
| $G_f$ | Fluid reasoning |
| $G_lr$ | Long-term storage and retrieval |
| $G_c$ | Comprehension-knowledge |
| $G_s$ | Cognitive processing speed |
| $G_v$ | Visual processing |
| $G_q$ | Quantitative knowledge |
| $G_a$ | Auditory processing |
| $G_rw$ | Reading and writing |
| $G_sm$ | Short-term memory |

be further divided as more detailed narrower capabilities. Noteworthy, the CHC model is not built upon the air; instead, it is quite a comprehensive structure that derived from large statistical data and several mathematic theories. "Factor Analysis" (FA) and "Item Response Theory" (IRT) are two primary adopted theories.

### B. Common Factor Model

In psychological research, factor analysis is one of the most important analytical methods on searching relationship between measurable quantities and intrinsic properties. Factor analysis is a direct extension of regression theory and partial correlation. A general form of factor analysis is:

$$\boldsymbol{y} = \boldsymbol{W}'\boldsymbol{x} + \boldsymbol{e} \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^n$ has $n$ observable dependent variables $y_1, \cdots, y_n$ with standard measurement in a population, and $m$ unobservable independent variables $x_1, \cdots, x_m$, known as *common factors* or *latent factors*; whereas $\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_n]$ (where $\boldsymbol{W} \in \mathbb{R}^{m\times n}$), known as "factor weights" contains $n$ vectors of regression weights that can be infer from $\boldsymbol{x}$ and $\boldsymbol{y}$; $\boldsymbol{e} \in \mathbb{R}^n$ is known as regression residual, which is uncorrelated with the independent variables, thus

$$\boldsymbol{E}\{\boldsymbol{x}\boldsymbol{e}\} = \boldsymbol{0} \tag{2}$$

By deploying common factor analysis, users may try to explain a series of different observable dependent variables by some independent latent factors, in recognition of empirical correlation matrices built by those observable variables. For example, measurements of students' performance on different subjects are usually obtained by tests. However, the performance is not absolutely unrelated to each other, which infers to performance correlations among different subjects. We can actually find how different subjects performance correlated with each other, or even further find some properties $\boldsymbol{x}$ of student that explain the observed correlations.

On the other hand, we can take advantage of the observed correlation to help designing more concise and comprehensive tests. *Spearman* attempted to explain those correlation via a global latent factor $g$, known as "general intelligence", to denote the ability that different performances reflect in common. Furthermore, since one common factor only reveals one communality that all tests share in common, it is reasonable to bring about the hypothesis that multivariate common factors model may reveal more communality among all tests.

Define $\boldsymbol{R}_x$ as the correlation matrix of the independent variables and $\boldsymbol{R}_y$ as the correlation matrix of the dependent variables. From Eq. (1), calculate correlation as

$$\boldsymbol{R}_y \ = \boldsymbol{W}'\boldsymbol{R}_x\boldsymbol{W} + \boldsymbol{E}^2 \tag{3}$$

where $\boldsymbol{E}^2$ is a diagonal matrix which stands for unique variances, and $\boldsymbol{R}$ is also a diagonal matrix because $\boldsymbol{x}$ are independent variables. Without further restriction to this equation, there should have infinitely many solutions [14]. Thus, denote $\boldsymbol{W}^* = \boldsymbol{Q}\boldsymbol{W}$, and $\boldsymbol{R}_x^* = (\boldsymbol{Q}^{-1})'\boldsymbol{R}_x\boldsymbol{Q}^{-1}$, which makes $(\boldsymbol{Q}^{-1})'\boldsymbol{R}_x\boldsymbol{Q}^{-1} = \boldsymbol{I}_m$. Eq. (3) can be written as

$$\boldsymbol{R}_y \ = (\boldsymbol{W}^*)'\boldsymbol{W}^* + \boldsymbol{E}^2 \tag{4}$$

Rewriting $\boldsymbol{R}_x$ to $\boldsymbol{R}_x^*$, can be explained as standardize deviation of each latent factors, which will adjust factor loadings correspondingly. Such a procedure is reasonable, because scaling on the factor loading won't affect the structure of common factor model. Therefore, the model parameters have been reduced to find factor loading of $\boldsymbol{W}^*$, and latent factors are restricted to be orthonormal.

When solving the problem of Eq. (4), infinitely many solutions of $\boldsymbol{W}^*$ can be found due to the unrestrictive dimension of $\boldsymbol{W}^*$ and "rotation" property of linear algebra. Practically to solve this common factor analysis problem, we can either (1) restrict the dimension of $\boldsymbol{W}^*$ under reasonable assumptions, which is called "confirmatory factor analysis" (CFA), or (2) tentatively set null hypothesis to accept dimension of latent factors as $m$, and calculate to identify if the null hypothesis is accepted or not; increase the dimension by one if failed to accept the null hypothesis until it being accepted; such a process is called exploratory factor analysis (EFA).

### C. Item Response Theory

In the CHC model, the stratum III defines a general ability $g$ to represent a general latent factor that reflects the communality all performances shared together. If we restrict the number of latent factors as one, the factor analysis model can be simplified to have one dimensional $\boldsymbol{W}$ and $x$. In this case, an item response model (IRM) is more applicable on exploring single latent trait of testers. A general unidimensional two-parameter dichotomous IRM is:

$$p_i(X = 1|\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}} \tag{5}$$

where $\theta$ represents latent trait level of tester, $a_i$, $b_i$ are two parameters of the model. $p_i(\theta)$ is shaping a cumulative distribution function (CDF), which represents the probability of observing a particular outcome of item $i$ by a tester with

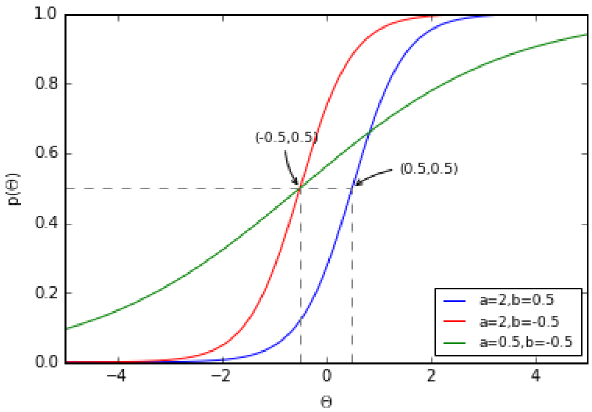


Figure 2: Two-parameter Item Response Model.

latent trait of $\theta$ (in this dichotomous model, the outcome is either $X = 0$ or $X = 1$).

Typical distributions of of $p_i(\theta)$ with different parameter settings are shown in Fig. 2. As illustrated by the figure, $b_i$ affects the location of $\theta$ where $p(X = 1|\theta) = 0.5$. It is often explained as difficulty of item $i$ that if the tester's latent trait level equals to the item difficulty, the probability of the tester observe the outcome $X = 1$ is $50\%$. $a_i$ is proportional to the slope of IRM curve, which affects the shape of the curve. It is recognized as discrimination parameter, the larger of $a$, the sheerer of the curve, thus the better of distinguishable capability of the item.

In this particular IRM, we restrict it to have two parameters, because more other parameters is considered rather unnecessary. And since we are inspecting the global cognitive capability $g$, or sometimes a particular broad cognitive capability $G_x$, an unidimensional is practically preferred considering complexity and workload. Dichotomous IRM is presented above; however, sometimes we get more than two category outcomes. For example, scale $[1, 2, 3, 4, 5]$, grade ["fail", "pass", "good", "excellent"], thus a polytomous IRM is expected. Detailed discussion is presented in section III.

## III. Theoretical Analysis

### A. Basic Rationale and Assumptions

When designing cognitive radio test batteries and applying performance evaluations, researchers are usually puzzled whether or not they should focus on recording the performance of the whole system or solely individual nodes. Ideally, it is highly desired to acquire performance indicators for the whole system with great details. However, as a potentially distributed or even hybrid system, CRNs can hardly be described using the performance globally. What's more difficult is that sometimes the structure of a CRN is usually dynamic and boundless. There always exist a dilemma whether to measure global network performance or partially radio device performance.

In our proposal, on the contrary, a more practical way is assumed to measure the performances of CRNs. It is to evaluate the performance of individual devices, and take the entire radio environment and network structure into consideration at the same time. In this section, let's clarify some important concepts in cognitive radio first, and then more theoretically discussions are presented on how to design test batteries and how to deploy them.

According to U.S. Federal Communications Commission (FCC), A CR is "a radio that can change its transmitter parameters based on interaction with the environment in which it operates. The majority of cognitive radios will probably be SDR but neither having software nor being field programmable are requirements of a cognitive radio" [19]. In this paper, however, a broader definition is adopted: "CR is a wireless device that is self-configurable and intelligently adaptive to environment".

A CRN is a communication network that composed by a set of interconnected CRs that can exchange data with each other. Since CR can hardly perform intelligence by itself as a single unit, CRN embraces many CRs to form a structural system in several forms, such as centralized, distributed or hybrid. Essentially, CRN is a form of cognitive systems.

A cognitive system is characterized by a cognition cycle, which consists of circler stages of observe, orient, plan, learn, decide and act [16]. Cognitive Radio System (CRS) is an even broader closure concept that considered not only the internal state of CRN, but also the output of the system. CRS is a dynamic system in which time plays a important role in its input-output behavior. Its cognitive behavior is comparable to human cognition due to the perception-action cycle, that *perceptor* will perceive the radio environment, and *actuator* take actions based on the processed information and feedback to *perceptor* to complete a cycle [12].

A CR is a complex radio unit embraced not only software modules but also hardware components. When evaluate "CR cognition", by definition, it considers to evaluate the complete structure of entire radio unit. While in reality, many CR research works considered one or two particular techniques instead of designing the complete CR. Therefore, a standard modularized CR structure is encouraged on evaluating one specific design of cognitive technique. With help of factor analysis, a large number of cognitive radios as well as a set of testing batteries are investigated. Thus we are able to construct a standard CR library that consists of some CR structures with replaceable modules with known cognitive capabilities. To evaluate a CR technique, simply replacing the module on standard CR with the new technique, and applying test batteries to inspect the new cognition capabilities and performance output.

Specifically, our proposed CR cognition evaluation methodology targets to:

- Construct standard test batteries that allows test-based evaluation on CRs, with weights coefficient of cognitive capabilities; and
- Optimize performance metrics in each standard test bat-

Table II: CR hardware setups.

| Hardware Parameters | Setup1 | Setup2 | Setup3 | Setup4 |
|---|---|---|---|---|
| Spectrum access range | DC∼10GHz | DC∼10GHz | DC∼3GHz | DC∼3GHz |
| Data Stream | up to 50 MS/s | up to 20 MS/s | up to 50 MS/s | up to 20 MS/s |
| Communication Range | 1000m | 100m | 1000m | 1000m |
| Power Supply | unlimited | unlimited | unlimited | unlimited |
| GPS equipment | Yes | No | No | No |

tery that embraces evaluation on every broad cognitive capabilities with minimal evaluation redundancy; and

- Construct standard reassemble-enabled CR models that allows flexible evaluation on CR techniques; and
- Exterminate standard test batteries to construct either application-oriented or general-purposed questionnaire that can measure the overall cognition.

### B. CR Physical Preparation

When modeling a CR structure or evaluating a CR technique, there are usually two options: either to simulate the CR via network simulators such as NS2, NS3, OmNET++, OPNET etc, or use SDR emulators such as BEE2 [7], SORA [2], USRP device [1], etc. Normally, network simulators allow more flexible hardware parameters adjustment, such as spectrum access range, computation capability, antenna setup, power supply, geographical information support etc; and network simulators allow extensive evaluation of CRs because they are highly scalable on modeling CRN size and topology. On the other hand, hardware-based SDR devices allow real-world implementation and evaluation of CR techniques, which produces more accurate performance result.

In the proposed CR evaluation model, we do not specify one type of CR hardware setup; instead, we allow a wide choice from the CR hardware configuration pool. Table II shows several examples of simulated CR hardware setups. For more convicing evaluation purpose, we are able to select from a wide choices among various SDR devices. In general, we are able to design a CR hardware configuration pool as $Pool_{hardware} = \{hd[1], hd[2], \cdots\}$. Our goal is to find distinctive and generally applicable hardware configurations to construct standard model evaluation library.

### C. CR Mental Capability

As discussed above, a CR is a wireless device with full stack of functionality design as shown in Fig. 3. In the physical layer, spectrum sensing is the process that the CR observes the radio environment and determines the status of spectrum resource via perceived data. Spectrum analysis/access is the process that the CR synthesizes the spectrum information in order to optimize communication by adaptively adjusting transmission parameters. Spectrum sharing/decision requires the CRs share available spectrum resources with each other, where a cognitive medium access control (MAC) protocol may be deployed. Adaptive routing allows a dynamic topology formation strategy for optimal spectrum efficiency and quality of service (QoS). Application layer is not an essential design of CR, thus some fundamental applications can be referred. In Table III, several well-known cognitive techniques are listed under different scopes, from which we are able to construct a set of full stack designed of CRs as $Pool_{stack} = \{st[1], st[2], \cdots\}$. We aim to locate several CRs being classified as “relatively” high cognition, average cognition and low cognition.

Table III: CR techniques in different layer.

| Different Layer | Techniques |
|---|---|
| Cognitive Engine | MARL [6]; Q-learning [11], etc |
| Spectrum Sensing | Energy detection; Match filter; Cyclostationary feature [3], etc |
| Spectrum Sharing/Decision | SenseLess [17]; Auction [13]; MAB based [4], etc |
| Network Layer | PoS [5], etc |
| App Layer | Video stream; Voice, etc |

### D. Test Battery

In the proposed test based CR evaluation model, test battery is essentially the most fundamental part of the methodology. A well designed test battery should contain features of:

- flexibility and scalability that allows ease of use by various CRs with different scale of CRN topology;
- comprehensiveness that inspect a wide range of comprehensive cognitive capabilities of CRs;
- conciseness that reduces redundant performance metrics for optimal evaluation; and
- versatileness that meets some specific requirements on CR evaluation.

Generally, similar to the CR physical preparation, there are two types of methods on designing test batteries. One is simulator-based event-driven test battery design, the other

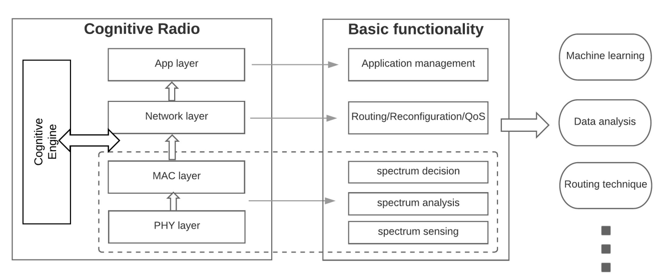


Figure 3: CR full stack structure.

is SDR emulator-based time-driven test battery design. The essential idea of the proposed CR evaluation model is to find the optimal design with proper platform instead of finding the best platform in general. As discussed earlier, our evaluation model does not restrict which way the test battery is designed. Considering CRs can be either deployed via SDR emulators or deployed by network simulators. Accordingly, a CR test battery can be designed in either way, however, we should discriminate between them for proper usage.

With emulator based time-driven design, radio environment is collected from the real world for direct usage or it can be further customized as different radio scenarios. For example, a combination of Wireless Open-Access Research Platform (WARP) and MATLAB is considered as emulator based time-driven CRN evaluation platform [18]. On the other hand, a simulated test battery requires concrete modeling of radio scenarios, such as radio propogation model, signal fluctuation, environment noise, etc. Denote different test battery scenario as $Pool_{scenario} = \{sc[1], sc[2], \cdots\}$.

## IV. CR Evaluation Model Design

The proposed CR evaluation model is not intended to test the actual performance of certain CR in specific scenarios. Instead, the model aims at measuring the latent cognitive capabilities of CRs that will offer references on evaluate the strength and weakness of CR techniques. In this section, a theoretical design is presented, which serves as the guideline of design CR evaluation models.

### A. Questionnaire Design

Although real world performance measurement is not the goal of our CR evaluation model, a proper set of performance metrics as the output of test battery is desired. However, distracted by massive performance metrics and different radio scenarios, which performance metric contributes to which cognitive capability remains undetermined.

Presumably, a linear relationship between test scenario performances and cognitive capabilities is defined as:

$$\boldsymbol{y} = \boldsymbol{W}'\boldsymbol{x} + \boldsymbol{e} \tag{6}$$

where $\boldsymbol{y}$ is the performance column vector of a CR:

$$cr[i,j] = \{hd[i] \in \{Pool_{hardware}\}, st[j] \in \{Pool_{stack}\}\}$$

tested in a test battery scenario: $sc[k] \in \{Pool_{scenario}\}$ (where $i, j, k = 0, 1, 2, \cdots$). $\boldsymbol{x}$ is the latent cognitive capability column vector of such CR, and $\boldsymbol{W}'$ is the weight coefficient matrix ($'$ indicates matrix transpose) of such test battery scenario which indicates the linear relation between performance metrics and cognitive capabilities. $\boldsymbol{e}$ defines the uncorrelated system noise.

Initially, we take a wide consideration of performance metrics from different aspects of CR, refer to table IV,

Table IV: CR performance metrics.

| Different Aspects | Performance Metrics |
|---|---|
| Perception | Primary User (PU) detection speed & rate, false alarm rate, etc |
| Learning | learning period, signal classification speed & accuracy, long term memory storage, short term memory storage & access speed, learning convergence time, radio adaptivity, etc |
| Action | communication range, spectrum usage efficiency, channel handoff speed, channel switch speed, adaptive resource distribution, etc |
| Network Layer | throughput, network scalability, packet error rate, packet routing cost, path reliability, mobility & adaptivity, etc |
| App Layer | Signal Noise Ratio (SNR), Quality of Service (QoS), power management, security, etc |

and acquire a high dimensional performance vector $\boldsymbol{y}$ with dimensionality of $p$. Apply test scenario $sc[k]$ to $n$ different CRs:$\{cr[i_1, j_1], cr[i_2, j_2], cr[i_3, j_3], \cdots, cr[i_n, j_n]\}$, we will construct a $p \times n$ performance matrix $\boldsymbol{Y}_{p\times n} = [\boldsymbol{y_1}, \boldsymbol{y_2}, \boldsymbol{y_3}, \cdots, \boldsymbol{y_n}]$. Accordingly, we will have a latent cognitive capability matrix $\boldsymbol{X}_{q\times n} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \cdots, \boldsymbol{x_n}]$, where $q$ is the number of latent cognitive capability factors, which unknown at current stage. Therefore, Eq. (6) can be further rewritten as:

$$\boldsymbol{Y} = \boldsymbol{W}'\boldsymbol{X} + \boldsymbol{E} \tag{7}$$

Eq. (7) will find the optimal solution of coefficient matrix $\boldsymbol{W}$ and latent factor matrix $\boldsymbol{X}$. Based on factor analysis theory, there are two reasonable assumptions to this model:

1) latent cognitive capabilities are uncorrelated to each other; and
2) latent cognitive capabilities are normalized, which can be written as $Cov(\boldsymbol{X}) = \boldsymbol{I}$.

Compute the covariance of Eq. (7) as:

$$\begin{aligned} Cov(\boldsymbol{Y}) &= \boldsymbol{W}'Cov(\boldsymbol{X})\boldsymbol{W} + Cov(\boldsymbol{E}) \\ &= \boldsymbol{W}'\boldsymbol{W} + Cov(\boldsymbol{E}) \end{aligned} \tag{8}$$

From Eq. (8), we are able to find a reasonable number of latent cognitive capabilities by selecting one of many criterion [14]. For example, principal component analysis (PCA) is one of commonly used FA algorithm, and usually the number of latent factors is determined by the number of eigenvalues larger than unit. The process is shown in Fig. 4. Further, apply orthogonal rotation on located coefficient matrix $W^*$, a rotated coefficient matrix $W^{**}$ is found to meet the coefficients convergence criterion. By analyzing rotated coefficient matrix $W^{**}$, we are able to eliminate some highly correlated performance metrics, because they are redundant on expressing latent cognitive capabilities. Consequently, a concise vector of performance metrics, that is well explanatory to CR cognitive capabilities, as well as a corresponding coefficient matrix $W^{**}$, is determined.

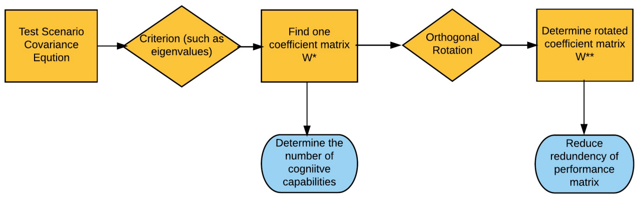


Figure 4: Factor Analysis of performance metrics on specific test scenario.

More generally, the wide choices of performance metrics can be applied to multiple test scenarios: $\{sc[k_1], sc[k_2], sc[k_3], \cdots\}$, which generates a concatenated performance matrix $\boldsymbol{\Upsilon}$:

$$\boldsymbol{\Upsilon} = \begin{bmatrix} \boldsymbol{Y_{k_1}} \\ \boldsymbol{Y_{k_2}} \\ \boldsymbol{Y_{k_3}} \\ \vdots \end{bmatrix}$$

Following the same process presented in Fig. 4, a set of consistent CR evaluation test batteries are constructed from multiple test scenarios. In addition, a set of CR examinees with known cognitive capabilities are determined. In consequence, two libraries are constructed:

1) a library of test batteries $TB = \{tb[1], tb[2], tb[3], \cdots\}$ that can be used to evaluate CR cognitive capabilities; and
2) a library of reassemble-enabled CR models $CRM = \{cr[i_1, j_1], cr[i_2, j_2], cr[i_3, j_3], \cdots\}$ that can be used to study single CR techniques.

### B. Questionnaire Examination

Usually, similar to human IQ test, users of CRs care more about the overall cognition of CRs or whether the CR is competent to a certain application scenario, instead of simply knowing the broad cognitive capabilities. For example, a user may want to find a CR that can work in a scarce-allocated non-rechargeable wireless sensor network (WSN), in which CR communication range, routing cost and power management maybe more critical than others. Another example is, a researcher may care about what is the generally intelligent of a CR, or what is the improvement gain from a CR technique. In former situation, user expects an application-oriented CR cognition notation, while researcher expects a general defined CR cognition notation. Either way can be expressed as:

$$\theta = \boldsymbol{c}'\boldsymbol{y} \tag{9}$$

where $\theta$ is the overall cognition of a CR, $\boldsymbol{c}$ is a coefficient vector which defines the importance scale of each measured performance metric from a test battery, and $\boldsymbol{y}$ is the concise performance metric vector. Compare Eq. (9) to Eq. (7), if we preset number of latent cognitive capability as one, and ignore the uncorrelated error, these two equations are basically identical except the reciprocal of coefficient vector. With different application requirement or general expectation, the coefficient vector can be assigned with different values for different purpose.

Because we expect to infer the overall cognition of CR from limited times of tests by using our designed test batteries, we apply polytomous IRM to construct our evaluation model.

Suppose there is a library of test batteries and a library of reassemble-enabled CR models (constructed from FA), and a preset coefficient (from application requirements), and then test different CRs with some test batteries for multiple times (e.g.1000 times), we are able to compute a set of parameters as the following:

1) Compute overall cognition $\theta_{ij}$ of each CR from average performances;
2) Normalize all CRs' cognition $\theta \sim (0, 1)$;
3) Classify the performance of each CR from each test in a response vector, noted as $class[1:4]$ =["fail", "pass", "good", "excellent"]; and
4) Compute the probability of certain response for each CR tested in each test battery $Pr(class[r]|_{\theta_{ij}, tb[k]}) = \frac{number\ of\ class[r]}{number\ of\ all\ tests}|_{\{theta_{ij}, tb[k]\}} \times 100\%$.

Refer to polytomous two parameter IRM, we model the test battery as:

$$Pr(class[r]|_{\theta_{ij}, tb[k]}) = \frac{1}{1 + e^{-a_k(\theta - b_{r,k})}} - \frac{1}{1 + e^{-a_k(\theta - b_{r-1,k})}} \tag{10}$$

From computed parameters of $\theta_{ij}$, $Pr(class[r]|_{\theta_{ij}, tb[k]})$, the polytomous IRM parameter $a_k$ and $b_{*,k}$ of each test battery can be estimated from Eq. (10). Figure 5 shows an example of test battery IRM that demonstrates the probability of observing a response by applying such test battery to CRs with different overall cognitions.

Finally, every test battery can be further designed for overall cognition measurement for either application-oriented or general-purposed basis.

### C. Questionnaire Application

Afterwards, with the construction of two libraries and specified examination, users can apply the evaluation model to evaluate:

1) Broad cognitive capabilities of a CR, which is analogical to stratum II in CHC model;
2) Cognitive capability improvement by a CR technique, when plugin such CR module to a CR from reassemble-enabled CR model library;
3) Overall cognition of a CR under certain criterion; and

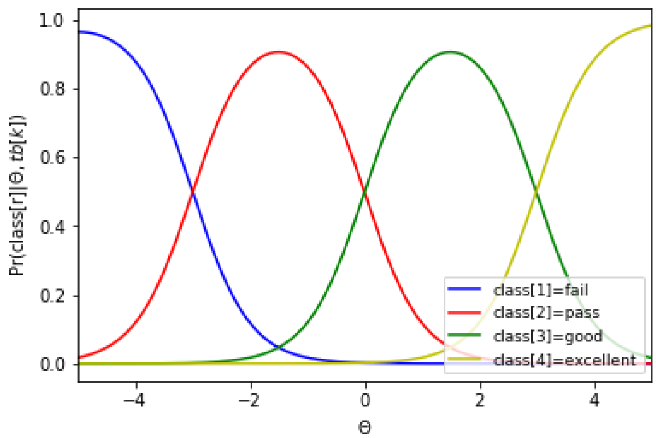


Figure 5: Probability of observing a response by applying a test battery to CRs with different overall cognitions.

4) Reference of the competence of a CR to an application.

Briefly speaking, the evaluation model shall offer the guidance to the strength and weakness of CR solutions, and provide advises on improvements.

## V. Conclusions and discussions

The CR techniques have been developed for year and the current CR systems are rather comprehensive and complex. It is highly desired but non-trivial to establish a common theoretical framework to guide the evaluation of current CR systems and provide insights for the further developments on CR techniques. The construction of a common factor model and questionnaires for CR systems is well discussed in this paper. However, there is a big gap between the abstract model and the real-world, diversified CR systems and CRN implementations. More practical deployments and investigations are needed by following the proposed steps, which may take great efforts.

There is an effort that shares the same vision but is focused on the intelligence measure of CRs in the MAC layer [8], in which the CHC based idea is applied to statistically study the IQ of CRs. This approach requires a comprehensive test and evaluation of as many as possible CR systems. The main challenges in constructing a cognition evaluation scheme that is able to be applied universally to all CR designs is the lack of sufficient samples. Following either the theoretical approach proposed in this work or the experimental plus statistical analysis oriented approaches, a comprehensive understanding of the design space of CRs or CRNs is mandatory. While what reported in this paper is far from mature, we hope this preliminary proposal could inspire more discussions and explorations to accelerate the research in this important area.

## References

[1] [Online]. Available: https://www.ettus.com/

[2] [Online]. Available: http://research.microsoft.com/en-us/projects/sora/

[3] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," in *Physical Communication*, 2011, pp. 40–62.

[4] A. B. H. Alaya-Feki, B. Sayrac, A. Le Cornec, and E. Moulines, "Semi dynamic parameter tuning for optimized opportunistic spectrum access," in *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*. IEEE, 2008, pp. 1–5.

[5] O. S. Badarneh and H. B. Salameh, "Probabilistic quality-aware routing in cognitive radio networks under dynamically varying spectrum opportunities," *Computers & Electrical Engineering*, vol. 38, no. 6, pp. 1731–1744 0045–7906, 2012.

[6] L. Busoniu, R. Babuska, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Transactions on Systems Man and Cybernetics Part C Applications and Reviews*, vol. 38, no. 2, pp. 156 1094–6977, 2008.

[7] C. Chang, J. Wawrzynek, and R. W. Brodersen, "Bee2: A high-end reconfigurable computing system," *IEEE Design and Test of Computers*, no. 2, 2005.

[8] M. Dabaghchian, A. Alipour-Fanid, S. Liu, K. Zeng, X. Liy, and Y. Chen, "Who is smarter? intelligence measure of learning-based cognitive radios," *arXiv preprint arXiv:1712.09315*, 2017.

[9] Q. Dong, Z. Yang, Y. Chen, K. Zeng, and X. Li, "Exploration of singular spectrum analysis for online anomaly detection in crns," in *EAI Endorsed Transactions on Security and Safety*, vol. 4, no. 12, 2017.

[10] D. P. Flanagan, V. C. Alfonso, and M. R. Reynolds, "Broad and narrow chc abilities measured and not measured by the wechsler scales: Moving beyond within-battery factor analysis," *Journal of Psychoeducational Assessment*, vol. 31, no. 2, pp. 202–223

[11] A. Galindo-Serrano and L. Giupponi, "Distributed q-learning for aggregated interference control in cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1823–1834

[12] S. Haykin, *Cognitive dynamic systems: perception-action cycle, radar and radio*. Cambridge University Press, 2012.

[13] S. K. Jayaweera, M. Bkassiny, and K. A. Avery, "Asymmetric cooperative communications based spectrum leasing via auctions in cognitive radio networks," *IEEE transactions on wireless communications*, vol. 10, no. 8, pp. 2716–2724

[14] R. P. McDonald, *Factor analysis and related methods*. Lawrence Erlbaum Associates, 1985.

[15] K. S. McGrew, "Chc theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research," *Intelligence*, vol. 37, no. 1, pp. 1–10

[16] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE personal communications*, vol. 6, no. 4, pp. 13–18 1070–9916, 1999.

[17] R. Murty, R. Chandra, T. Moscibroda, and P. V. Bahl, "Senseless: A database-driven white spaces network," in *IEEE TRANSACTIONS ON MOBILE COMPUTING*, 2012.

[18] J. J. Thompson, K. M. Hopkinson, and M. D. Silvius, "A test methodology for evaluating cognitive radio systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6311–6324

[19] A. M. Wyglinski, M. Nekovee, and T. . . Hou, *Cognitive radio communications and networks: principles and practice*. Academic Press, 2009.

[20] Y. Zhao, S. Mao, J. O. Neel, and J. H. Reed, "Performance evaluation of cognitive radios: Metrics, utility functions, and methodology," *Proceedings of the IEEE*, vol. 97, no. 4, pp. 642–659 2009.