Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 51-34 (2019) 53–60

Optimal Decision Fusion Under Order Effects ⋆

Ehsan Nekouei * **Jieqiang Wei** * **John S. Baras** **,* **Mikael Skoglund** *
Karl H. Johansson *

* *KTH Royal Institute of Technology, Stockholm, Sweden (e-mail: {nekouei,jieqiang,skoglund, kallej}@kth.se).*
** *University of Maryland, College Park (e-mail:baras@isr.umd.edu)*

Abstract: This paper studies an optimal decision fusion problem with a group of human decision makers when an *order effect* is present. The order effect refers to situations wherein the process of decision making by a human is affected by the order of decisions. In our set-up, all human decision makers, called observers, receive the same data which is generated by a common but unknown hypothesis. Then, each observer independently generates a sequence of decisions which are modeled by employing non-commutative probabilistic models of the data and their relation to the unknown hypothesis. The use of non-commutative probability models is motivated by recent psychological studies which indicate that these non-commutative probability models are more suitable for capturing the order effect in human decision making, compared with the classical probability model. A central decision maker (CDM) receives (possibly a subset of) the observers' decisions and decides on the true hypothesis. The considered problem becomes an optimal decision fusion problem with observations modeled using a non-commutative (Von Neumann) probability model. The structure of the optimal decision rule at the CDM is studied under two scenarios. In the first scenario, the CDM receives the entire history of the observers' decisions whereas in the second scenario, the CDM receives only the last decision of each observer. The perfromance of the optimal fusion rule is numerically evaluated and compared with the optimal fusion rule derived when using a classical probability model.

Keywords:

1. INTRODUCTION

1.1 Motivation

Human intelligence is widely employed in crowd-based hypothesis testing problems, *e.g.,* in crowd-sourcing applications such as Amazon Mechanical Turk. In these applications, a group of human beings receive common data related to an unknown hypothesis, and *independently* generate decisions (regarding the validity or not of the hypothesis) based on these received data. Then, a *fusion* algorithm is used to combine the observers' decisions and form a reliable outcome for the hypothesis testing problem. The performance of this framework depends on: (i) the decision fusion algorithm, (ii) the models of the behavior of human decision makers, hereafter called observers. Decision fusion algorithms are typically designed based on simple probabilistic models of the observer's behavior, *e.g.,* see Ok et al. (2016).

However, recent psychological studies show that non-commutative probability models may be more suitable, compared with classical probability models, for explaining the human judgment process, for example when the *order effect* is present, see Busemeyer et al. (2015) and references therein. The order effect implies that the order, in which the judgments are made, influences our decision making process. This is due to the fact that making a judgment influences our cognition system which in turn influences our next judgment. This observation motivates us to study the optimal decision fusion algorithm in a hypothesis testing problem with multiple human observers when the order effect is present. To further clarify the order effect in human judgment, we present the following two well-known examples from the psychology literature.

The first example is the primacy-recency effect in a medical inference task studied in Bergus et al. (1998). In this study, 315 physicians were asked to diagnose the existence of a certain disease in a patient based on the patient's history and physical examination along with the clinical test data. The physicians were divided into two groups. In one group, the physicians first received the patient's history whereas in the other group the physicians first received the examination and clinical test data. The authors found that the probability of correct diagnostic is substantially different among the two groups, signifying the order effect in the associated human decision making.

The second example examines the order effect in a jury task studied in Trueblood and Busemeyer (2011) using multiple experiments. In one experiment, 299 university students were asked to serve as jurors for a hypothetical criminal trial where they were presented with a prosecution and a defense. Then, they were asked whether the defendant is guilty or not. The authors varied the order and strength of the prosecution and the defense among participants. The statistical analysis of this experiment shows that the participants' decisions depend on the order in which the prosecution and defense were presented. The authors also studied the perfromance of the (non-commutative) Von Neumann probability model in capturing the order effect in the above experiments as well as other experiments including the jury task experiments in McKenzie et al. (2002). Their

⋆ This work is supported by the Knut and Alice Wallenberg Foundation, the Swedish Foundation for Strategic Research and the Swedish Research Council. This work is also supported by USA ARO grant W911NF-15-1-0646

2405-8963
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2019.01.022

results show that the Von Neumann probability model provides a better fit to the empirical data when the order effect is present, compared with several classical models including the belief adjustment and the minimum acceptable strength models.

Motivated by these results, in the current paper, the decision making process by human observers is modeled using the Von Neumann probability model. Intuitively, this model can be viewed as a stochastic belief evolution model wherein the current belief of a decision maker affects her decision, then her belief is updated based on her current and past decisions.

1.2 Order Effect in Von Neumann Probability Model

The order effect is a key distinguishing point between non-commutative probability models and the classical probability model developed by Andrey Kolmogorov. More precisely, in the classical model, the order in which two random variables are generated does not affect their joint distribution since classical distributions operate on measurable sets using set operations. These operations form a (commutative) Boolean algebra. However, the joint distribution of two (sequentially generated) random variables employing non-commutative probability models depends on their generation order. To pictorially illustrate this point, let X and Y denote two random variables which are generated sequentially using the Von Neumann model and take values in $\{1, 2, 3, 4\}$ (see next section for more details on the Von Neumann model). Fig. 1(a) shows the trajectories of the empirical probability of the event $X = 1, Y = 2$ under different generation orders of X and Y. According to this figure, the trajectories of the empirical probability converge to two distinct values. This observation is in accordance with the fact that the joint probability of the events depends on their order of generation.

One may speculate that it is possible to utilize the notion of conditional probability to express the statistical properties of X and Y using a pair of classical random variables. However, this approach violates Bayes' law in classical probability. To numerically illustrate this point, we generated samples from the above example under different generation orders of X and Y. Then, we empirically estimated $\Pr(Y = 2 \,|X = 1)$ ($\Pr(X = 1 \,|Y = 2)$) using the samples in which X (Y) was generated first. Fig. 1(b) shows the empirical values of $\Pr(X = 1 \,|Y = 2) \Pr(Y = 2)$ and that of $\Pr(Y = 2 \,|X = 1) \Pr(X = 1)$ as a function of the number of samples. As this figure shows, the classical probability model constructed in this way violates Bayes' law. Finally, it is worth nothing that *there does not exist* a one-to-one correspondence, from a Von Neumann model with dimension more than two to the classical model, which satisfies certain desirable statistical properties, *e.g.,* conservation law (functional subordination), see Section 1.4. and Proposition 1.4.1. in Holevo (2001).

Remark 1. In the rest of the paper, we use "$;$" to denote the order of generation of random variables, *i.e.,* $(X = 1; Y = 2)$ implies Y is generated first.

Remark 2. We note that non-commutative probability models are alternative frameworks to the classical model. They range from the quantum probability model developed by John Von Neumann, Von Neumann (1955), to the free probability, a highly non-commutative model, developed by Dan Voiculescu, Voiculescu et al. (1992). These models have a diverse application domain, *e.g.,* studying the physical interactions in atomic scale, quantum communications Hayashi (2016) and characterizing the asymptotic behavior of eigenvalues of random matrices Tao (2012). In this paper, by a "non-commutative" model (*e.g.,* in Fig. 1(a)), we refer to the Von Neumann model which is usually applied to the micro-world.

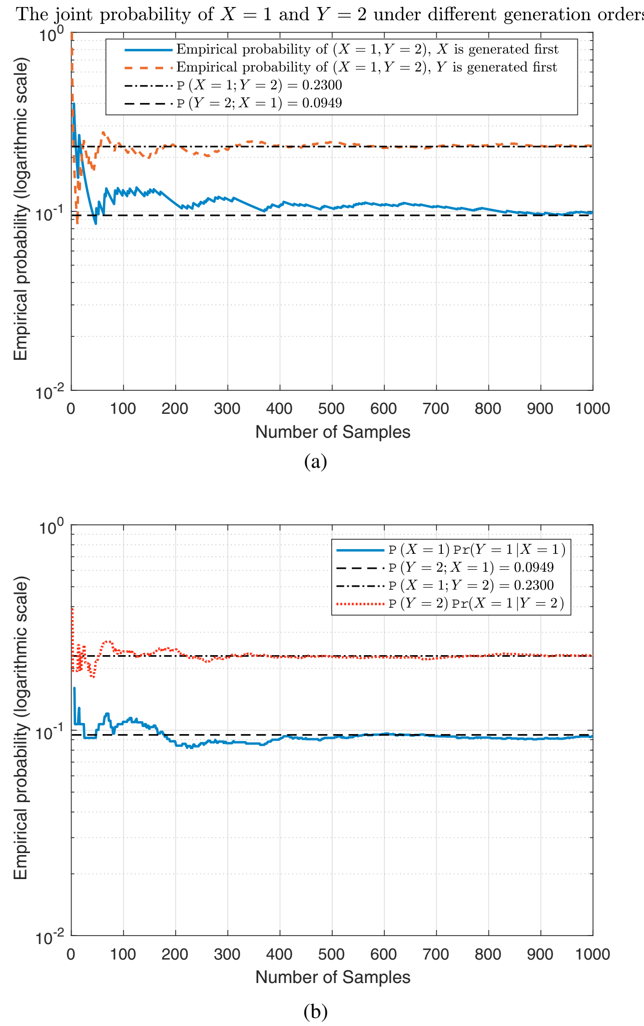


Fig. 1. The empirical probability of the event $X = 1$ and $Y = 2$ in the Von Neumann probability model for different generation orders (a), the empirical estimates of $\Pr(X = 1 \,|Y = 2) \Pr(Y = 2)$ and $\Pr(Y = 2 \,|X = 1) \Pr(X = 1)$ (b) as a function of the number of samples.

1.3 Related Work

The optimal detection rule for a single quantum system has been extensively investigated in the literature, *e.g.,* see Holevo (2001), Helstrom (1969) and Baras (1987). The error exponent of the optimal decision rule, jointly performed on a number of identical quantum systems, in the limit of large number of systems was studied in Nussbaum and Szkoła (2009). The authors in Baras et al. (1976) considered the problem of minimum variance filtering of a scalar signal from quantum mechanical measurements and derived the necessary and sufficient conditions for jointly optimal measurement operators and post-processing matrices. These results were extended to vector valued signals in Baras and Harger (1977) and Baras (1988). The interested reader is refereed to Hayashi (2016) for a complete treatment of asymptotics in quantum hypothesis testing problems.

In the psychology literature, the Von Neumann model has been used to explain the order effect in human decision making experiments. We note that such connections were discussed in a much earlier paper by Baras, Baras (1979). The authors in Busemeyer and Wang (2017) considered two sequential decision making experiments. In the first experiment, the respondents answered the same question sequentially, AA experiment, whereas in the second experiment, they answered three questions wherein the first and the last questions were the same, ABA experiment. The correlation analysis in Busemeyer and Wang (2017) shows that the Von Neumann model is more suitable for capturing the order effect compared with the classical model. The authors in Wang et al. (2014) proposed a statistical quantity, named quantum question equality, to test the suitability of the Von Neumann model for explaining the order effect. Their empirical studies, using surveys and laboratory experiments, supports the Von Neumann model for explaining the order effect in human decision making experimental data.

The authors in Boyer-Kassem et al. (2016) proposed the notion of the grand reciprocity as a statistical test to verify the suitability of the Von Neumann model in describing order effect in human decision making. Their empirical analysis shows that degenerate Von Neumann models provide a better fit to the experimental data. The paper Aerts and de Bianchi (2017) proposed a model, called general tension reduction, to capture both the order effect and response replicability in human decision making process. The authors in Kellen et al. (2017) derived a classical model to capture the order effect by introducing auxiliary random variables. However, these random variables are not observable from empirical data, *e.g.,* a binary random variable which indicates whether the first decision of an observer affects her second decision or not.

1.4 Contributions

In this paper, we study the optimal decision fusion problem in a hypothesis testing problem with N human observers who receive common data related to an unknown hypothesis. Then, each observer independently generates a sequence of decisions. To capture the order effect, the observers' decisions are derived based on the non-commutative Von Neumann model. A central decision maker (CDM) receives possibly a subset of the observers' decisions. The problem of optimal decision fusion at the CDM is studied under two scenarios. In the first scenario, the CDM has access to the entire history of the observers' decisions while in the second scenario, the CDM only has access to the last decisions of the observers. The performance of the optimal decision is numerically evaluated and compared with the optimal fusion rule derived using the classical probability model.

This paper is structured as follows. Next section describes the considered hypothesis testing problem as well as the non-commutative probability model for the decision making of human observers. The optimal decision fusion rules at the CDM are derived in Section 3. Our numerical results are presented in Section 4 followed by the concluding remarks in Section 5. The proofs are relegated to appendices to aid readability.

2. SYSTEM MODEL

Consider a hypothesis testing problem with L hypotheses and N human observers. All observers are exposed to the same data

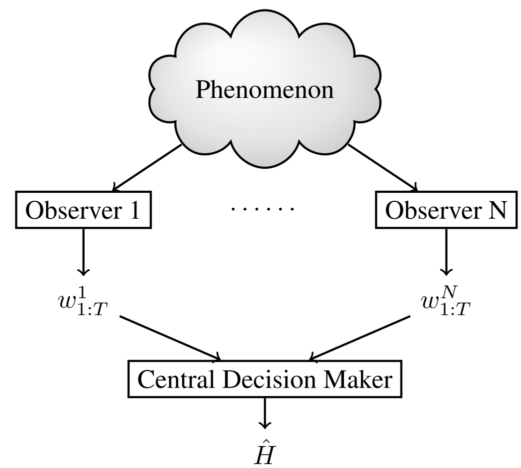


Fig. 2. A hypothesis testing problem with N observers over T time steps.

which are related to one of the L hypotheses $\{H_1, \cdots, H_L\}$, *e.g.,* a picture. Let p_i denote the occurrence probability of hypothesis H_i. Each observer makes T sequential decisions according to the available data, *e.g.,* she answers a set of multiple choice questions related to the observed information. In our set-up, a central decision maker (CDM) at time $T+1$ receives the (possibly a subset of) observers' decisions and forms its estimate of the underlying hypothesis.

Let $\boldsymbol{W}_{1:k} = \left(W^1_{1:k}, \cdots, W^N_{1:k}\right)$ denote the collection of the observers' decisions up to time k, where $W^j_{1:k}$ represents the decisions of observer j up to time k. Two different decision making scenarios are considered at the CDM. In the first scenario, each observer generates a sequence of T decisions and transmits its sequence of decisions to the CDM at time $T+1$. Thus, the CDM has access to $\boldsymbol{W}_{1:T}$ at time $T+1$. Under the second scenario, each observer makes T decisions, but only transmits her final decision to the CDM. That is, the CDM receives $\boldsymbol{W}_T$ at time $T+1$ and does not have access to the history of the decisions made by observers. The interaction between the CDM and observers is pictorially depicted in Fig. 2.

2.1 Non-commutative Decision Making Model

The behavior of each observer under H_i is modeled using the finite dimensional Von Neumann probability model. In this model, the density operator ρ^i_1, $i \in \{1, \cdots, L\}$, plays the role of a classical probability distribution (or belief) (typically of the state of a system or agent). The density operator ρ^i is an $n \times n$, positive definite Hermitian matrix with

$$\mathsf{Tr}\left(\rho^i_1\right) = 1$$

where $\mathsf{Tr}\left(\cdot\right)$ denotes the trace operator. In our set-up, ρ^i_1 represents the initial belief or a-priori bias of observers when they are exposed to the observed (perceived) data related to hypothesis H_i.

The outcome of the decision making process of each observer is modeled by a positive operator-valued measure (POVM). Let $\Omega_k = \{1, \cdots, l_k\}$ encode the set of possible decisions of each observer at time k and let M_k denote the POVM associated with the decision made at time k, *i.e.,* $M_k = \{M_k(\omega)\}_{w\in\Omega_k}$, consists of l_k $(n \times n)$ Hermitian positive definite matrices with

$$\sum_{w\in\Omega_k} M_k(w) = I$$

where I is the identity matrix. Here, w represents a certain decision made by an observer and $M_k(w)$ is used to compute the probability of the event w as it will be described in the next paragraph.

We next explain how the probabilities of different events are computed. Let w_k^j denote a decision of observer j at time k and $w_{1:k}^j$ denote a sequence of decisions of observer j from time 1 to k. We use $\rho_{k+1}^i\left(w_{1:k}^j\right)$ to represent the belief of observer j at time $k+1$, *i.e.,* the density operator associated with observer j at time $k+1$, under H_i. The dependency of $\rho_{k+1}^i\left(w_{1:k}^j\right)$ on $w_{1:k}^j$ reflects the fact that each observer is biased according to her previous decisions. Given $\rho_{k+1}^i\left(w_{1:k}^j\right)$, the probability of selecting $w_{k+1}^j \in \Omega_{k+1}$ at time $k+1$ can be written as

$$\mathsf{Pr}\left(W_{k+1}^j = w_{k+1}^j \left|\rho_{k+1}^i\left(w_{1:k}^j\right)\right.\right) = \mathsf{Tr}\left(M_{k+1}\left(w_{k+1}^j\right)\rho_{k+1}^i\left(w_{1:k}^j\right)\right) \quad (1)$$

The density operator $\rho_{k+1}^i\left(w_{1:k}^j\right)$ is computed recursively according to

$$\rho_{k+1}^i\left(w_{1:k}^j\right) = \frac{M_k^{\frac{1}{2}}\left(w_k^j\right)\rho_k^i\left(w_{1:k-1}^j\right)M_k^{\frac{1}{2}}\left(w_k^j\right)}{\mathsf{Tr}\left(M_k\left(w_k^j\right)\rho_k^i\left(w_{1:k-1}^i\right)\right)} \quad (2)$$

with

$$\rho_2^i\left(w_{1:1}^j\right) = \frac{M_1^{\frac{1}{2}}\left(w_1^j\right)\rho_1^i M_1^{\frac{1}{2}}\left(w_1^j\right)}{\mathsf{Tr}\left(M_1\left(w_1^j\right)\rho_1^i\right)} \quad (3)$$

Based on (2), the belief of the observer j at time $k+1$ depends on the entire history of her decisions. Moreover, her belief affects her decision at time $k+1$ according to (1). Thus, this decision making model can be viewed as a stochastic non-commutative belief evolution model.

2.2 *Composite Systems*

Assume that the decision making processes of observers are independent of each other. Then, the joint behavior of observers can be described by a composite system. Let $\mathcal{H}_i$ denote the composite system of N observers under the hypothesis H_i. Then, the density operator of $\mathcal{H}_i$ at time k can be written as the following tensor product

$$\boldsymbol{\rho}_k^i(\boldsymbol{w}_{1:k-1}) = \rho_k^i\left(w_{1:k-1}^1\right)\otimes\cdots\otimes\rho_k^i\left(w_{1:k-1}^N\right)$$

and the measurement operator at time k is described by

$$M_k = \left\{M_k\left(w_k^1\right)\otimes\cdots\otimes M_k\left(w_k^N\right)\right\}_{w_k^j\in\Omega_k,\forall j}$$

The POVM associated with the joint outcome $\boldsymbol{w}_k = \left(w_k^1,\cdots,w_k^N\right)$ is denoted by $M_k(\boldsymbol{w}_k) = M_k\left(w_k^1\right)\otimes\cdots\otimes M_k\left(w_k^N\right)$.

3. MAIN RESULTS

In this section, we study the structure of the optimal decision fusion rule when (a) the CDM receives the entire history of the observers' decisions and (b) when it receives the last decisions of observers. We start by presenting the optimal decision fusion rule when the entire history is available at the CDM.

3.1 *Optimal Decision Using The Entire History of Decisions*

In this subsection, we first derive the structure of the optimal decision fusion rule at the CDM. To this end, consider the randomized decision rule in which, given $\boldsymbol{W}_{1:T} = \boldsymbol{w}_{1:T}$, the CDM selects H_i with probability $t_i(\boldsymbol{w}_{1:T})$ where $\sum_{i=1}^L t_i(\boldsymbol{w}_{1:T}) = 1$ for all $\boldsymbol{w}_{1:T} \in \Omega_1^N\times\cdots\times\Omega_T^N$. Let $\mathsf{Pr}(C)$ denote the probability of correct decision at the CDM. Then, $\mathsf{Pr}(C)$ can be written as

$$\begin{aligned}\mathsf{Pr}(C) &= \mathsf{E}\left[\mathsf{Pr}(C\,|\boldsymbol{W}_{1:T})\right]\\ &= \sum_{\boldsymbol{w}_{1:T}}\sum_{i=1}^L t_i(\boldsymbol{w}_{1:T})\,\mathsf{Pr}(\boldsymbol{W}_{1:T} = \boldsymbol{w}_{1:T}\,|H_i)\,p_i\end{aligned}$$

Note that $\mathsf{Pr}(C)$ is a function of the $t_i(\cdot)$'s. Thus, the optimal decision fusion rule at the CDM, which maximizes $\mathsf{Pr}(C)$, is the solution of the following optimization problem

$$\begin{aligned}\max_{\{t_i(\boldsymbol{w}_{1:T})\}_{i,\boldsymbol{w}_{1:T}}} &\sum_{\boldsymbol{w}_{1:T}}\sum_{i=1}^L t_i(\boldsymbol{w}_{1:T})\,\mathsf{Pr}(\boldsymbol{W}_{1:T} = \boldsymbol{w}_{1:T}\,|H_i)\,p_i\\ &\sum_{i=1}^L t_i(\boldsymbol{w}_{1:T}) = 1, \quad \forall \boldsymbol{w}_{1:T}\\ &0 \le t_i(\boldsymbol{w}_{1:T}) \quad \forall i, \boldsymbol{w}_{1:T}\end{aligned}$$

This optimization problem can be written as a set of decoupled optimization problems. Thus, solving the above optimization problem is equivalent to solving the following optimization problem

$$\begin{aligned}\max_{\{t_i(\boldsymbol{w}_{1:T})\}_i} &\sum_{i-1}^L t_i(\boldsymbol{w}_{1:T})\,\mathsf{Pr}(\boldsymbol{W}_{1:T} = \boldsymbol{w}_{1:T}\,|H_i)\,p_i\\ &\sum_{i=1}^L t_i(\boldsymbol{w}_{1:T}) = 1\\ &0 \le t_i(\boldsymbol{w}_{1:T}) \quad \forall i\end{aligned}$$

for each $\boldsymbol{w}_{1:T} \in \Omega_1^N\times\cdots\times\Omega_T^N$. It can be easily shown that the solution of the above optimization problem is given by $t_i(\boldsymbol{w}_{1:T}) = 1$ if $i = \arg\max_i p_i\mathsf{Pr}(\boldsymbol{W}_{1:T} = \boldsymbol{w}_{1:T}\,|H_i)$.

Next lemma formally states the structure of the optimal decision fusion law at the CDM.

Lemma 1. Consider the decision fusion problem at the CDM with the entire history of decisions. Then, the optimal fusion rule is given by the solution of the following optimization problem

$$\arg\max_i p_i\mathsf{Pr}(\boldsymbol{W}_{1:T} = \boldsymbol{w}_{1:T}\,|H_i)$$

According to Lemma 1 the optimal decision fusion rule based on the entire history of decision, *i.e.,* $\boldsymbol{w}_{1:T}$, is to select the hypothesis with maximum likelihood of occurrence given $\boldsymbol{w}_{1:T}$. To solve the optimization problem above, we need to compute the joint probability of the event $\boldsymbol{W}_{1:T} = \boldsymbol{w}_{1:T}$ given H_i for all i.

In the next lemma, a formula is derived for computing $\Pr(\boldsymbol{W}_{1:T} = \boldsymbol{w}_{1:T} | H_i)$.

Lemma 2. The conditional probability of observing $\boldsymbol{w}_{1:T}$ under H_i is given by

$$\Pr(\boldsymbol{W}_{1:T} = \boldsymbol{w}_{1:T} | H_i) = \prod_{j=1}^{N}\prod_{k=1}^{T} \mathrm{Tr}\left(M_k\left(w_k^j\right)\rho_{k-1}^i\left(w_{1:k-1}^j\right)\right) \quad (4)$$

where $\rho_{k-1}^i\left(w_{1:k-1}^j\right)$ satisfies the recursion (2).

Proof. See Appendix A

The optimal decision fusion rule at the CDM, given $\boldsymbol{w}_{1:T}$, is obtained by combing Lemmas 1 and 2. This result is presented in the next theorem.

Theorem 1. Assume that the entire history of the observers' decisions is available at the CDM. Then, the optimal fusion rule at the CDM is given by

$$\arg\max_i p_i \prod_{j=1}^{N}\prod_{k=1}^{T} \mathrm{Tr}\left(M_k\left(w_k^j\right)\rho_{k-1}^i\left(w_{1:k-1}^j\right)\right) \quad (5)$$

where $\rho_{k-1}^i\left(w_{1:k-1}^j\right)$ satisfies the recursion (2).

3.2 Optimal Decision Rule Using The Final Decisions

In this subsection, we derive the optimal decision fusion rule at the CDM when only the last decisions are available at the CDM. To this end, let $\boldsymbol{W}_T$ denote the final decisions of the observers and $\boldsymbol{w}_T$ be one of its realizations. Similar to the derivation of Lemma 1, it can be easily shown that the optimal decision rule, given $\boldsymbol{w}_T$, can be written as

$$\arg\max_i p_i \Pr(\boldsymbol{W}_T = \boldsymbol{w}_T | H_i) \quad (6)$$

The next lemma derives a recursive expression for computing the $\Pr(\boldsymbol{W}_T = \boldsymbol{w}_T | H_i)$.

Lemma 3. The probability $\Pr(\boldsymbol{W}_T = \boldsymbol{w}_T | H_i)$ can be expressed as

$$\Pr(\boldsymbol{W}_T = \boldsymbol{w}_T | H_i) = \mathrm{Tr}\left(M_T(\boldsymbol{w}_T)\boldsymbol{\rho}_T^i\right)$$

where $\boldsymbol{\rho}_T^i = \rho_T^i \otimes \cdots \otimes \rho_T^i$ and ρ_T^i is computed using the recursion

$$\rho_{k+1}^i = \sum_{l_k} M_k^{\frac{1}{2}}(l_k)\,\rho_k^i M_k^{\frac{1}{2}}(l_k) \quad (7)$$

with $k \geq 1$.

Proof. See Appendix B.

According to Lemma 3, the density operator for evaluating $\Pr(\boldsymbol{W}_T = \boldsymbol{w}_T | H_i)$ is computed forward in time using (7). Here, ρ_k^i is a density operator representing the effect of prior beliefs of an observer regarding hypothesis H_i at time k.

The next theorem presents the optimal fusion rule at the CDM when only the final decisions are available.

Theorem 2. Assume that the CDM only has access to the last decision of each observer. Then, the optimal fusion rule is given by

$$\arg\max_i p_i \mathrm{Tr}\left(M_T(\boldsymbol{w}_T)\boldsymbol{\rho}_T^i\right) \quad (8)$$

where $\boldsymbol{\rho}_T^i = \rho_T^i \otimes \cdots \otimes \rho_T^i$ and ρ_T^i is computed by propagating the initial density under H_i forward in time using (7).

To implement the optimal fusion rules, the CDM requires the knowledge of $\left\{\rho_1^i\right\}_i$, $\{p_i\}_i$, $\{M_k\}_k$ as well as the decisions of observers. Moreover, the computational complexity of the optimal fusion rule in (5) is at most $O\left(LTNn^3\right)$ and that of (8) is at most $O\left(LNn^3 + Ln^3\sum_k |\Omega_k|\right)$.

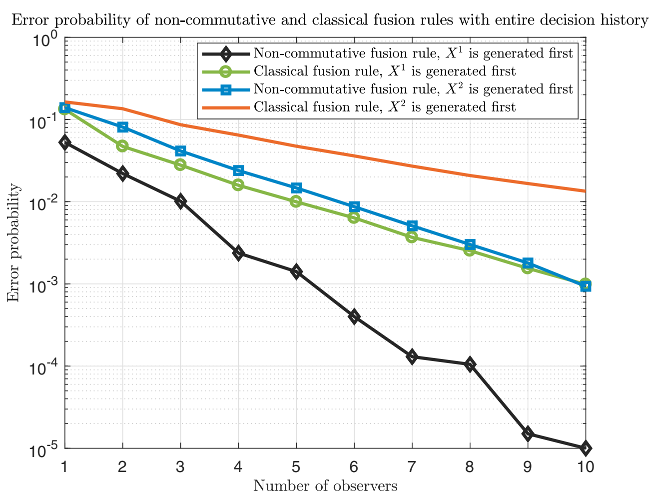
4. NUMERICAL RESULTS

In this section, we consider a binary hypothesis testing problem and numerically evaluate the performance of the optimal fusion rule, in terms of the error probability, when the CDM has access to the whole history of decisions and when it has access to the last decisions. We also compare the perfromance of the fusion rule, derived from the non-commutative probability model, with that of the optimal decision fusion rule obtained from the classical probability model. A two-stage decision making problem is considered wherein the observers' decisions are modeled by two random variables X^1 and X^2 from the Von Neumann probability model. In our model, X^1 and X^2 take values in $\{1,2,3,4\}$ and the associated POVMs are assumed to be projection valued measures. The density operators are two rank-one Hermitian positive definite matrices. It is assumed that the two hypotheses H_1 and H_2 are equally likely. In our numerical results, the density operators and the POVMs are randomly selected.

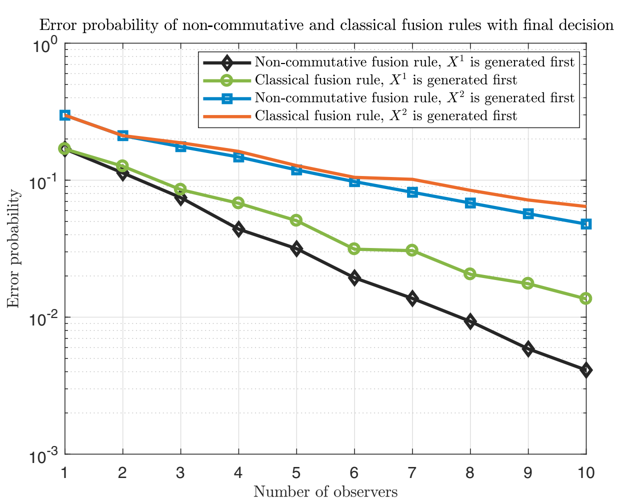
Let X_c^1 and X_c^2 denote two classical random variables taking values in $\{1,2,3,4\}$. Next, we construct a classical joint probability distribution for X_c^1 and X_c^2, under each hypothesis, from empirical observations of X^1 and X^2. The constructed distributions are used to obtain the optimal classical decision fusion rule. To this end, independent copies (samples) of X^1 and X^2 are generated under different orders and hypotheses using the density evolution rule (2). Let $D_i = D_i\left(x^2; x^1\right) \cup D_i\left(x^1; x^2\right)$ ($i \in \{0,1\}$) denote the collected samples under H_i. Here, $D_i\left(x^2; x^1\right)$ is the set of samples where X^1 is generated first and $D_i\left(x^1; x^2\right)$ is the set of samples where X^2 is generated first. In our numerical results, the classical model is obtained by discarding the order from data, *i.e.,* the joint distribution of X_c^1 and X_c^2 under H_i is generated using D_i. This construction of the classical probability model respects the Bayes' law. We note that it is possible to obtain an empirical conditional probability distribution of X_c^1 given X_c^2 under H_i using $D_i\left(x^1; x^2\right)$ and that of X_c^2 given X_c^1 using $D_i\left(x^2; x^1\right)$. However, it is straightforward to verify that the classical probability model formed by these empirical conditional distributions will not respect Bayes' law due to the non-commutativity of the Von Neumann model (for more details, see Fig. 1(b) and its description in Subsection 1.2).

Fig. 3(a) shows the error probability of the classical and the non-commutative fusion rules, under different decision making orders, when the entire history of decisions is available at the CDM. According to this figure, the optimal non-commutative fusion rule always outperforms the optimal classical fusion rule. Moreover, the performance gap between the non-commutative and the classical fusion rules increases as the number of observers becomes large. This observation indicates that the non-commutative fusion rule is more efficient in incorporating the new information provided by new observers in order to improve its perfromance. Based on Fig. 3(a), the perfromance of the optimal fusion rule depends on the generation order of X^1 and X^2. This is due to the facts that the optimal fusion rule depends on the beliefs of observers, and the belief of an observer evolves differently under different generation orders.

Fig. 3(b) shows the error probability of the classical and non-commutative fusion rules when only the last decisions of the observers are available at the CDM. As this figure shows, the non-commutative fusion rule achieves a lower error probability compared with the classical fusion rule. However, the performance of the CDM heavily degrades when only the final decisions are available at the CDM since it has access to a limited amount of information compared with the scenario where it has access to the complete history of decisions.



(a)



(b)

Fig. 3. The error probability at CDM under the non-commutative and classical fusion rules versus the number of decision makers with (a) entire history, (b) final decisions.

At this point, one might be tempted to attribute the order effect to an auxiliary random variable. Let α denote a binary random variable where $\alpha = 1$ denotes X^1 is generated first and $\alpha = 2$ denotes X^2 is generated first. Thus, we have $\left(X^2; X^1\right) \iff \left(X^1, X^2, \alpha = 1\right)$ and $\left(X^1; X^2\right) \iff \left(X^2, X^1, \alpha = 2\right)$. The joint distribution of X^1 and X^2 conditioned on $\alpha = 1$ and H_i can be estimated using $D_i\left(x^2; x^1\right)$. Similarly, the joint distribution of X^1 and X^2 conditioned on $\alpha = 2$ under H_i can be estimated using $D_i\left(x^1; x^2\right)$. Also the conditional distribution of α can be obtained as $\Pr\left(\alpha = 1 \,|H_i\right) = \frac{\left|D_i\left(x^2;x^1\right)\right|}{|D_i|}$ and $\Pr\left(\alpha = 1 \,|H_i\right) = \frac{\left|D_i\left(x^2;x^1\right)\right|}{|D_i|}$. Note that the conditional distribution of α is purely a function of the number of samples in the sets $D_i\left(x^2; x^1\right), D_i\left(x^1; x^2\right)$ $i = 0, 1$, rather than the statistical properties of data, which might degrade the perfromance of the classical estimator. We refer the reader to Trueblood and Busemeyer (2011) for a more detailed discussion on this approach.

5. CONCLUSIONS

In this paper, we studied the optimal decision fusion problem with a group of human observers when the order effect has been manifested. The decision making process of observers was modeled using the Von Neumann probability model. The optimal decision fusion rule at a central decision maker (CDM) was studied under two main scenarios. In the first scenario, the CDM receives the entire history of the decisions made by the observers whereas in the second scenario, the CDM only knows the last decision of each observer. The performance of the optimal fusion rule was numerically studied.

REFERENCES

Aerts, D. and de Bianchi, M.S. (2017). Beyond-quantum modeling of question order effects and response replicability in psychological measurements. *Journal of Mathematical Psychology*, 79, 104 – 120.

Baras, J., Harger, R., and Park, Y. (1976). Quantum-mechanical linear filtering of random signal sequences. *IEEE Transactions on Information Theory*, 22(1), 59–64.

Baras, J.S. (1979). Noncommutative probability models in quantum communication and multi-agent stochastic control. *Richerche di Automatica*, (2), 217–264.

Baras, J.S. (1987). Distributed asynchonous detection: General models. In *26th IEEE Conference on Decision and Control*, volume 26, 1832–1835.

Baras, J.S. and Harger, R. (1977). Quantum mechanical linear filtering of vector signal processes. *IEEE Transactions on Information Theory*, 23(6), 683–693.

Baras, J.S. (1988). An optimization problem from linear filtering with quantum measurements. *Applied Mathematics and Optimization*, 18(1), 191–214.

Bergus, G., Chapman, G., T. Levy, B., W. Ely, J., and Oppliger, R. (1998). Clinical diagnosis and the order of information. 18, 412–7.

Boyer-Kassem, T., Duchêne, S., and Guerci, E. (2016). Testing quantum-like models of judgment for question order effect. *Mathematical Social Sciences*, 80, 33 – 46.

Busemeyer, J. and Wang, Z. (2017). Is there a problem with quantum models of psychological measurements? *PLoS ONE*.

Busemeyer, J., Wang, Z., Townsend, J., and Eidels, A. (2015). *Oxford Handbook of Computational and Mathematical Psychology*. Oxford Library of Psychology Series. Oxford University Press, Incorporated.

Gudder, S. (2007). Quantum probability. In *Handbook of Quantum Logic and Quantum Structures*, 121 – 146. Elsevier Science B.V., Amsterdam.

Hayashi, M. (2016). *Quantum Information Theory: Mathematical Foundation*. Graduate Texts in Physics. Springer Berlin Heidelberg.
Helstrom, C.W. (1969). Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2), 231–252.
Holevo, A. (2001). *Statistical Structure of Quantum Theory*. Lecture Notes in Physics Monographs. Springer.
Kellen, D., Singmann, H., and Batchelder, W.H. (2017). Classic-probability accounts of mirrored (quantum-like) order effects in human judgments. *Decision*, 2325–9965.
McKenzie, C.R.M., Lee, S.M., and Chen, K.K. (2002). When negative evidence increases confidence: change in belief after hearing two sides of a dispute. *Journal of Behavioral Decision Making*, 15(1), 1–18.
Nussbaum, M. and Szkoła, A. (2009). The chernoff lower bound for symmetric quantum hypothesis testing. *The Annals of Statistics*, 37(2), 1040–1057.
Ok, J., Oh, S., Shin, J., and Yi, Y. (2016). Optimality of belief propagation for crowdsourced classification. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, 535–544. JMLR.org.
Tao, T. (2012). *Topics in Random Matrix Theory*. Graduate studies in mathematics. American Mathematical Soc.
Trueblood, J.S. and Busemeyer, J.R. (2011). A quantum probability account of order effects in inference. *Cognitive Science*, 35(8), 1518–1552.
Voiculescu, D., Dykema, K., and Nica, A. (1992). *Free Random Variables*. CRM Monograph Series. American Mathematical Soc.
Von Neumann, J. (1955). *Mathematical Foundations of Quantum Mechanics*. Investigations in physics. Princeton University Press.
Wang, Z., Solloway, T., Shiffrin, R.M., and Busemeyer, J.R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences*, 111(26), 9431–9436.

Appendix A. PROOF OF LEMMA 2

Note that for $k \geq 2$, $\mathsf{Pr}\left(\boldsymbol{W}_{1:k} = \boldsymbol{w}_{1:k} \left| H_i \right.\right)$ can be written as

$$\mathsf{Pr}\left(\boldsymbol{W}_{1:k} = \boldsymbol{w}_{1:k} \left| H_i \right.\right) = \mathsf{Tr}\left(\mathcal{M}_k\left(\boldsymbol{w}_k\right)\left[\mathcal{M}_{k-1}\left(\boldsymbol{w}_{k-1}\right)\left[\cdots \mathcal{M}_1\left(\boldsymbol{w}_1\right)\left[\boldsymbol{\rho}_1^i\right]\right]\right]\right) \quad \text{(A.1)}$$

where $\mathcal{M}\left(\cdot\right)\left[\cdot\right]$ is called an "instrument" Holevo (2001), Baras (1987), Baras (1988). Given a density operator ρ and a measurement operator M, $\mathcal{M}\left(M\left(\omega\right)\right)\left[\rho\right]$ is defined as the probability of observing w as the outcome of measurement multiplied by the posterior density operator after observing w, *i.e.,*

$$\mathcal{M}\left(w\right)\left[\rho\right] = \mathsf{Tr}\left(M\left(w\right)\rho\right)\rho\left(w\right) \quad \text{(A.2)}$$

and $\rho\left(w\right)$ is the posterior density operator after observing w which is given by Gudder (2007)

$$\rho\left(w\right) = \frac{M^{\frac{1}{2}}\left(w\right)\rho M^{\frac{1}{2}}\left(w\right)}{\mathsf{Tr}\left(M\left(w\right)\rho\right)} \quad \text{(A.3)}$$

Next lemma studies the structure of the instrument $\mathcal{M}\left(\cdot\right)\left[\cdot\right]$ in our problem.

Lemma 4. The instrument $\mathcal{M}_k\left(\cdot\right)\left[\cdot\right]$ satisfies

$$\mathcal{M}_k\left(\boldsymbol{w}_k\right)\left[\mathcal{M}_{k-1}\left(\boldsymbol{w}_{k-1}\right)\left[\cdots \mathcal{M}_1\left(\boldsymbol{w}_1\right)\left[\boldsymbol{\rho}_1^i\right]\right]\right] = \boldsymbol{\rho}_{k+1}^i\left(\boldsymbol{w}_{1:k}\right)\prod_{j=1}^{N}\prod_{t=1}^{k}\mathsf{Tr}\left(M_t\left(w_t^j\right)\rho_t^i\left(w_{1:t-1}^j\right)\right) \quad \text{(A.4)}$$

where $\rho_1^i\left(w_{1:0}^j\right) = \rho_1^i$ and $\boldsymbol{\rho}_1^i = \underbrace{\rho_1^i \otimes \cdots \otimes \rho_1^i}_{N\text{folds}}$

Proof. See Appendix C.1.

The desired result follows from the above lemma and the fact that $\mathsf{Tr}\left(\boldsymbol{\rho}_{T+1}^i\left(\boldsymbol{w}_{1:T}\right)\right) = 1$ as $\boldsymbol{\rho}_{k+1}^i\left(\boldsymbol{w}_{1:k}\right)$ is a density operator.

Appendix B. PROOF OF LEMMA 3

Note that $\mathsf{Pr}\left(\boldsymbol{W}_k = \boldsymbol{w}_k \left| H_i \right.\right)$ can be written as

$$\begin{aligned}
&\mathsf{Pr}\left(\boldsymbol{W}_T = \boldsymbol{w}_T \left| H_i \right.\right) \\
&= \mathsf{Pr}\left(\boldsymbol{W}_T = \boldsymbol{w}_T; \boldsymbol{W}_{T-1} \in \Omega_{T-1}^N; \ldots; \boldsymbol{W}_1 \in \Omega_1^N \left| H_i \right.\right) \\
&= \mathsf{Tr}\left(\mathcal{M}_T\left(\boldsymbol{w}_T^j\right)\left[\mathcal{M}_{T-1}\left(\Omega_{T-1}^N\right)\left[\cdots \mathcal{M}_1\left(\Omega_1^N\right)\left[\boldsymbol{\rho}_1^i\right]\right]\right]\right)
\end{aligned} \quad \text{(B.1)}$$

where $\Omega_k^N = \underbrace{\Omega_k \times \cdots \times \Omega_k}_{N\text{folds}}$, Ω_k is the probability space at time k. For $k = 1$, we have

$$\begin{aligned}
\mathcal{M}_1\left(\Omega_1^N\right)\left[\boldsymbol{\rho}_1^i\right] &\overset{(a)}{=} \sum_{\boldsymbol{w}_1} M_1^{\frac{1}{2}}\left(\boldsymbol{w}_1\right)\boldsymbol{\rho}_1^i M_1^{\frac{1}{2}}\left(\boldsymbol{w}_1\right) \\
&\overset{(b)}{=} \left(\sum_{w_1^1} M_1^{\frac{1}{2}}\left(w_1^1\right)\rho_1^i M_1^{\frac{1}{2}}\left(w_1^1\right)\right) \otimes \cdots \\
&\quad \otimes \left(\sum_{w_1^N} M_1^{\frac{1}{2}}\left(w_1^N\right)\rho_1^i M_1^{\frac{1}{2}}\left(w_1^N\right)\right) \\
&= \rho_2^i \otimes \cdots \otimes \rho_2^i \\
&= \boldsymbol{\rho}_2^i
\end{aligned}$$

where (a) follows from the additive property of the instrument (see property 3 in Holevo (2001) page 97) and (b) follows from the fact that tensor product is a multi-linear map. Now, assume that we have

$$\begin{aligned}
\mathcal{M}_k\left(\Omega_k^N\right)\left[\cdots \mathcal{M}_1\left(\Omega_1^N\right)\left[\boldsymbol{\rho}_1^i\right]\right] &= \boldsymbol{\rho}_{k+1}^i \\
&= \rho_{k+1}^i \otimes \cdots \otimes \rho_{k+1}^i
\end{aligned}$$

Then, we have

$$\begin{aligned}
&\mathcal{M}_{k+1}\left(\Omega_{k+1}^N\right)\left[\cdots \mathcal{M}_1\left(\Omega_1^N\right)\left[\boldsymbol{\rho}_1^i\right]\right] \\
&\quad = \sum_{\boldsymbol{w}_{k+1}} M_{k+1}^{\frac{1}{2}}\left(\boldsymbol{w}_{k+1}\right)\boldsymbol{\rho}_{k+1}^i M_{k+1}^{\frac{1}{2}}\left(\boldsymbol{w}_{k+1}\right) \\
&\quad = \rho_{k+2}^i \otimes \cdots \otimes \rho_{k+2}^i \\
&\quad = \boldsymbol{\rho}_{k+2}^i
\end{aligned} \quad \text{(B.5)}$$

Using (B.1) and (B.5), we have

$$\mathsf{Pr}\left(\boldsymbol{W}_T = \boldsymbol{w}_T \left| H_i \right.\right) = \mathsf{Tr}\left(M_T\left(\boldsymbol{w}_T\right)\boldsymbol{\rho}_T^i\right)$$

Appendix C. PROOF OF AUXILIARY LEMMAS

C.1 Proof of Lemma 4

To prove this result, we study the structure of the density operator at time k under H_i in the next lemma.

$$
\begin{aligned}
&\mathcal{M}_{k+1}\left(\boldsymbol{w}_{k+1}\right)\left[\mathcal{M}_{k}\left(\boldsymbol{w}_{k}\right)\left[\cdots \mathcal{M}_{1}\left(\boldsymbol{w}_{1}\right)\left[\boldsymbol{\rho}^{i}\right]\right]\right] \\
&\quad=\rho_{k+2}^{i}\left(\boldsymbol{w}_{1:k+1}\right) \operatorname{Tr}\left(M_{k+1}\left(\boldsymbol{w}_{k+1}\right) \mathcal{M}_{k}\left(\boldsymbol{w}_{k}\right)\left[\cdots \mathcal{M}_{1}\left(\boldsymbol{w}_{1}\right)\left[\boldsymbol{\rho}^{i}\right]\right]\right) \\
&\quad=\rho_{k+2}^{i}\left(\boldsymbol{w}_{k+1}\right)\left(\prod_{j=1}^{N} \prod_{t=1}^{k} \operatorname{Tr}\left(M_{t}\left(w_{t}^{j}\right) \rho_{t}^{i}\left(w_{1:t-1}^{j}\right)\right)\right) \operatorname{Tr}\left(M_{k+1}\left(\boldsymbol{w}_{k+1}\right) \boldsymbol{\rho}_{k+1}^{i}\left(\boldsymbol{w}_{1:k}\right)\right) \\
&\quad=\rho_{k+2}^{i}\left(\boldsymbol{w}_{k+1}\right)\left(\prod_{j=1}^{N} \prod_{t=1}^{k} \operatorname{Tr}\left(M_{t}\left(w_{t}^{j}\right) \rho_{t}^{i}\left(w_{1:t-1}^{j}\right)\right)\right) \prod_{j} \operatorname{Tr}\left(M_{k+1}\left(w_{k+1}^{j}\right) \rho_{k+1}^{i}\left(w_{1:k}^{j}\right)\right)
\end{aligned}
\tag{B.2}
$$

$$
\begin{aligned}
\boldsymbol{\rho}_{2}^{i}\left(\boldsymbol{w}_{1}\right) &= \frac{M_{1}^{\frac{1}{2}}\left(\boldsymbol{w}_{1}\right) \boldsymbol{\rho}_{1}^{i} M_{1}^{\frac{1}{2}}\left(\boldsymbol{w}_{1}\right)}{\operatorname{Tr}\left(M_{1}\left(\boldsymbol{w}_{1}\right) \boldsymbol{\rho}_{1}^{i}\right)} \\
&\overset{(a)}{=} \frac{\left(M_{1}^{\frac{1}{2}}\left(w_{1}^{1}\right) \rho_{1}^{i} M_{1}^{\frac{1}{2}}\left(w_{1}^{1}\right)\right) \otimes \cdots \otimes\left(M_{1}^{\frac{1}{2}}\left(w_{1}^{N}\right) \rho_{1}^{i} M_{1}^{\frac{1}{2}}\left(w_{1}^{N}\right)\right)}{\operatorname{Tr}\left(M_{1}\left(w_{1}^{1}\right) \rho_{1}^{i}\right) \times \cdots \times \operatorname{Tr}\left(M_{1}\left(w_{1}^{N}\right) \rho_{1}^{i}\right)} \\
&= \rho_{2}^{i}\left(w_{1}^{1}\right) \otimes \cdots \otimes \rho_{2}^{i}\left(w_{1}^{N}\right)
\end{aligned}
\tag{B.3}
$$

$$
\begin{aligned}
\boldsymbol{\rho}_{k+1}^{i}\left(\boldsymbol{w}_{1:k}\right) &= \frac{M_{k}^{\frac{1}{2}}\left(\boldsymbol{w}_{k}\right) \boldsymbol{\rho}_{k}^{i}\left(\boldsymbol{w}_{1:k-1}\right) M_{k}^{\frac{1}{2}}\left(\boldsymbol{w}_{k}\right)}{\operatorname{Tr}\left(M_{k}\left(\boldsymbol{w}_{k}\right) \boldsymbol{\rho}_{\boldsymbol{w}_{1:k-1}}^{i}\right)} \\
&= \frac{\left(M_{k}^{\frac{1}{2}}\left(w_{k}^{1}\right) \rho_{k}^{i}\left(w_{1:k-1}^{1}\right) M_{k}^{\frac{1}{2}}\left(w_{k}^{1}\right)\right) \otimes \cdots \otimes\left(M_{k}^{\frac{1}{2}}\left(w_{k}^{N}\right) \rho_{k}^{i}\left(w_{1:k-1}^{N}\right) M_{k}^{\frac{1}{2}}\left(w_{k}^{N}\right)\right)}{\operatorname{Tr}\left(M_{k}\left(w_{k}^{1}\right) \rho_{k}^{i}\left(w_{1:k-1}^{1}\right)\right) \times \cdots \times \operatorname{Tr}\left(M_{k}\left(w_{k}^{N}\right) \rho_{k}^{i}\left(w_{1:k-1}^{1}\right)\right)}
\end{aligned}
\tag{B.4}
$$

Lemma 5. The density operator at time k, $\boldsymbol{\rho}_k^i(\boldsymbol{w}_{1:k-1})$, can be written as

$$\boldsymbol{\rho}_{k}^{i}\left(\boldsymbol{w}_{1:k-1}\right)=\rho_{k}^{i}\left(w_{1:k-1}^{1}\right) \otimes \cdots \otimes \rho_{k}^{i}\left(w_{1:k-1}^{N}\right) \tag{C.1}$$

where $\rho_k^i(\cdot)$ is obtained from the recursion (2).

Proof. See Appendix C.2.

.We prove this result by induction. For $k=1$, we have

$$
\begin{aligned}
\mathcal{M}_{1}\left(\boldsymbol{w}_{1}\right)\left[\boldsymbol{\rho}^{i}\right] &= \operatorname{Tr}\left(M_{1}\left(\boldsymbol{w}_{1}\right) \boldsymbol{\rho}_{1}^{i}\right) \boldsymbol{\rho}_{2}^{i}\left(\boldsymbol{w}_{1}\right) \\
&\overset{(a)}{=} \boldsymbol{\rho}_{2}^{i}\left(\boldsymbol{w}_{1}\right) \prod_{j} \operatorname{Tr}\left(M_{1}\left(w_{1}^{j}\right) \rho_{1}^{i}\right)
\end{aligned}
$$

where (a) follows from previous lemma. Assume that for at time k, (A.4) holds. Then, at time $k+1$, we have (B.2).

The desired result follows from A.1 and Lemma 5.

C.2 Proof of Lemma 5

We first show this result for $k=2$. Note that the initial density of the composite system under H_i can be written as $\boldsymbol{\rho}_1^i=\rho_1^i \otimes \cdots \otimes \rho_1^i$. Also, the measurement operator corresponding to $\boldsymbol{w}_1$ can be written as $M_1(\boldsymbol{w}_1)=M_1(w_1^1)\otimes\cdots\otimes M_1(w_1^N)$. Using (A.3), $\boldsymbol{\rho}_2^i(\boldsymbol{w}_1)$ can be written as (B.3) where (a) follows from the fact that $M_1^{\frac{1}{2}}(\boldsymbol{w}_1)=M_1^{\frac{1}{2}}(w_1^1)\otimes\cdots\otimes M_1^{\frac{1}{2}}(w_1^N)$. Now assume that $\boldsymbol{\rho}_k^i(\boldsymbol{w}_{1:k-1})$ is given by (C.1). Using (A.3) $\boldsymbol{\rho}_{k+1}^i(\boldsymbol{w}_{1:k})$ can be expressed as (B.4).