Universidade de São Paulo
Instituto de Física

Mecânica Estatística de Sistemas de Agentes Bayesianos

Aplicação à Teoria dos Fundamentos Morais

Jônatas Eduardo da Silva César

Orientador:
Prof. Dr. Nestor Felipe Caticha Alfonso

Comissão Examinadora:

Prof. Dr. Nestor Felipe Caticha Alfonso (IF-USP)
Prof[a]. Dr[a]. Carmen Pimentel Cintra do Prado (IF-USP)
Prof. Dr. Marcelo Martinelli (IF-USP)
Prof. Dr. José Raimundo Novaes Chiappin (FEA-USP)
Prof. Dr. Osame Kinouchi Filho (FFCLRP-USP)

Tese de doutorado apresentada ao Instituto de Física para a obtenção do título de Doutor em Ciências.

São Paulo
2014

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação do Instituto de Física da Universidade de São Paulo

César, Jônatas Eduardo da Silva

Mecânica estatística de sistemas de agentes Bayesianos: aplicação à teoria dos fundamentos morais. São Paulo, 2014.

Tese (Doutorado) – Universidade de São Paulo. Instituto de Física. Departamento de Física Feral

Orientador: Prof. Dr. Nestor Felipe Caticha Alfonso

Área de Concentração: Física

Unitermos: 1. Mecânica estatística; 2. Redes neurais; 3.Psicologia social; 4. Simulação (Estatística); 5. Simulação de sistemas.

USP/IF/SBI-013/2014

The Three Theorems of Psychohistorical Quantitivity:
The population under scrutiny is oblivious to the existence of the science of Psychohistory.
The time periods dealt with are in the region of 3 generations.
The population must be in the billions ($\pm$75 billions) for a statistical probability to have a psychohistorical validity.

—-Fundation, Isaac Azimov, 1920-1992

Agradecimentos

A cada ano que passa eu percebo que o trabalho científico depende de muitas outras coisas além da capacidade intelectual do cientista. Certamente, para desempenharmos bem as tarefas que nós escolhemos e nos tornamos melhores homens precisamos estar cercados de pessoas que admiramos e que podemos contar em momentos de necessidades e fartura. Posso dizer com muita certeza que os bons resultados que obtive em meu doutorado só foram possíveis graças a participação essencial de diversas pessoas.

Primeiramente, como não poderia deixar de ser, devo agradecer à minha família, minha mãe Ana Lúcia da Silva César, meu pai Eduardo Antonio César, minha irmã mais velha Janaina da Silva César e meu irmão caçula Paulo da Silva César, aos meus avós paternos José Augusto e Maria Abadia; e maternos João e Regina. Agradeço à minha esposa Layra Saori Okimura César, com quem celebrei matrimonio recentemente, também agradeço à minha sogra Teresa Okimura, à cunhada Denise Lie Okimura e ao cunhado Leonardo Junichi Okimura. Vocês são importantes para mim de tantas maneiras que seria impossível para mim redigir o quão grato eu sou pelas inumeráveis maneiras que vocês me ajudaram. Por isso, eu vou dizer apenas muito obrigado, vocês sempre podem contar comigo da mesma maneira que eu sei posso contar com vocês.

Em segundo lugar, quero agradecer aos meu amigos, pessoas que deixaram minha vida muito mais rica, pessoas que compartilhei momentos de seriedade e bebedeira. Muito obrigado Alexandre Patriota, Felippe Alexandre Barbosa, Juliano Neves, Álvaro Diego, Henrrique Miguel, João Bosco, Rodrigo Alves de Lima, Alexandre André dos

Santos, Mateus Jensen Didonet, Marcelo Boareto. Vocês fizeram que os anos do meu doutoramento fossem memoráveis. Espero que continuemos a criar juntos boas memórias mesmo no anos mais distantes de nossos futuro.

Também devo os meus mais sinceros agradecimentos ao meu orientador, Nestor Caticha, por ter me dado liberdade e apoio nas horas em que eu precisei além da orientação em um projeto que inicialmente parecia, para mim, não muito promissor, mas que no fim do meu doutorado mostrou-se ser um trabalho que deve continuar valendo o meu esforço e de muitas outras pessoas que irão trabalhar nele. Também agradeço aos meus colegas de grupo pela convivência estimulante e agradável, por toda a ajudas que vocês me deram e que fizeram com que meu doutorado fosse possível de ser finalizado. Muito obrigado, João Pedro, André Manuel, Renato Vicente, Rafael Calsaverine, André Maizel, Eduardo Dubai, Bruno Golfette, Guilherme, Edgar, Diogro, Felippe e Bruno.

Por fim, agradeço à Fundação de Amparo a Pesquisa do Estado de São Paulo - FAPESP pelo apoio financeiro.

Resumo

Moral e ideologia política estão intrinsecamente relacionados com aprendizado, tipos de personalidade e estratégias cognitivas de indivíduos. Usando um modelo de agentes Bayesianos adaptativos e interagentes tentaremos responder como características do aprendizado moral na infância e adolescência estão relacionadas à ideologia, traços de personalidade e estratégias cognitivas. Assumimos que o aprendizado moral do agente pode ser dividido em duas fases. A primeira fase é uma mímica do aprendizado de pessoas na infância e adolescência. Nessa fase, o modelo se assemelha ao aprendizado Bayesiano supervisionado, onde a estratégia para lidar com novas informações muda com a quantidade de informação recebida. Posteriormente, na segunda fase, agentes com estratégias cognitivas fixas discutem assuntos públicos, com conteúdo moral, e mudam suas opiniões com a motivação de diminuir o custo psicológico de discordância com seus parceiros sociais.

Comparando as assinaturas estatísticas das opiniões dos agentes na segunda fase com assinaturas similares obtidas através do Questionário dos Fundamentos Morais, concluímos que nosso modelo apresenta diversas caraterísticas que tem respaldo experimental. Por exemplo, a quantidade de informação moral julgada na primeira fase está positivamente correlacionada com o liberalismo. Além disso, agentes que são estatisticamente identificados como liberais se adaptam mais rapidamente a mudanças na sociedade. Também constatamos que com o aumento do parâmetro de nosso modelo denominado pressão social, agentes estatisticamente identificados com pessoas liberais passam a ter perfis estatísticos mais parecidos com os de pessoas conservadores.

Os métodos usados neste estudo, simuações de Monte Carlo, aproximação de campo médio, são típicos da Mecânica Estatistica.

Abstract

Moral and political ideology are intrinsically related with learning processes, personality traits and individual cognitive strategies. Using an adaptive interacting Bayesian agent model we try to understand how characteristics of childhood and adolescent moral learning are related with ideology, personality traits, and cognitive strategies. We assume that the agent's moral learning can be divided in two phases. The first phase is a mimic of the learning processes of individuals in childhood and adolescence, in this phase, the model resembles the Bayesian supervised learning, where the strategy to deal with new information changes with the total amount of received information. Later, in the second phase, agents with frozen cognitive strategies discuss public issues, with moral content, and change its opinion motivated to decrease the psychological cost of disagreement with its social partners.

Comparing the statistical signatures of agent's opinions in the second phase with similar signatures obtained from data of the Moral Foundations Theory Questionnaire, we conclude that our model presents several features that have experimental support. For example, the amount of moral information acquired in the first phase is positively correlated with liberalism. Moreover, agents which are statistically identified as liberal adapt more quickly to changes in society. We also found that with increase of the social pressure parameter, agents statistically identified as liberals will have statistical profiles more similar with conservatives.

The methods used in this study, Monte Carlo simulations, mean field approximation, are typical of Statistical Mechanics.

Prólogo

Toda pesquisa científica tem suas peculiaridades. Em trabalhos multidisciplinares existe uma grande dificuldade de sintetizar o conteúdo de áreas de conhecimentos com histórias e *modus operandi* completamente distintos. Este trabalho tem o objetivo de descrever o aprendizado moral usando técnicas matemáticas de Mecânica Estatística e Inferência Bayesiana. O texto é escrito para um leitor que tenha pouca familiaridade tanto com os métodos matemáticos quanto com as teorias de ciência social apresentadas. Para isso, o conteúdo necessário para entender o trabalho é apresentado em duas partes. Na primeira parte, destacamos as teorias e experimentos relacionados ao aprendizado moral, dando pouca ênfase aos detalhes matemáticos de nosso modelo. Esperamos, com isso, que um leitor com pouco domínio das técnicas matemáticas usadas seja capaz de entender a essência de nossa modelagem. A segunda parte do texto é uma série de apêndices, sendo os dois primeiros escritos para que um leitor interessado pelos detalhes matemáticos seja capaz de entender melhor algumas características do modelo. O terceiro e último apêndice apresenta uma série de resultados para o modelo matemático em que baseamos o texto principal, escritos sob o contexto de dinâmicas de opiniões.

Conteúdo

⌈1⌉ Introdução

> Let us apply to the political and moral sciences the method founded on observation and calculation, the method with has served us so well in the natural sciences.
>
> — Pierre-Simon Laplace, 1749-1827

A Física atual atingiu um grau de sofisticação e controle impressionante. Com arcabouço matemático relativamente consistente, nós somos capazes de entender ou prever com grande precisão o comportamento de eventos e objetos de escalas subatômicas até escalas cosmológicas. Isso se torna ainda mais impressionante quando comparamos esse poder de descrição do *universo físico* com a nossa baixa capacidade de predição do comportamento humano.

Além da grande variabilidade e complexidade dos fenômenos sociais, um dos motivos da dificuldade de compreensão dos mesmos são os nossos vieses cognitivos. Entre eles, a tendencia de achar que fenômenos sociais são simples de entender. Por exemplo: no livro *Tudo é Óbvio*[1] o sociólogo Duncan Watts informa ao leitor que durante a segunda guerra mundial soldados de origem rural apresentavam melhores índices de bem-estar quando comparados com os de origem urbana. De fato, essa informação parece bem plausível já que é fácil de imaginar que na terceira década do século passado a vida em áreas rurais era muito mais complicada do que em áreas urbanas o que possibilitaria uma melhor adaptação dessas pessoas durante a guerra. No entanto, o grupo de pessoas que realmente reportaram melhores índices de bem-estar foram os de origem urbana; note que essa informação também é bastante plausível já que é fácil de imaginar que pessoas de

[1] D. J. Watts. *Tudo É Óbvio - Desde Que Você Saiba a Resposta*. Paz e Terra, São Paulo, 2011

origem urbana estão mais adaptadas em viver em espaços pequenos e com alta concentração e também estão mais acostumadas a lidar com vários níveis de hierarquia. Com esse argumento, vemos que, de fato, os fenômenos sociais devem ser estudados cuidadosamente, e principalmente levando em conta dados experimentais, já que argumentos de bom senso nem sempre revelam a natureza do fenômeno de forma acurada.

Até o fim do último século os fenômenos sociais recebiam pouca atenção da comunidade de Física; no entanto, com o grande poder de processamento dos computadores modernos, diversos modelos simples de dinâmica coletiva foram propostos a partir de princípios de senso comum, para modelar alguns fenômenos sociais. De fato, a aplicação de métodos de Mecânica Estatística está se disseminando nas mais variadas áreas do conhecimento. Esses métodos são reconhecidamente bem-sucedidos na descrição de fenômenos físicos; com eles, propriedades macroscópicas são descritas quando temos modelos do funcionamento dos constituintes microscópicos do sistema em estudo. Uma revisão da área é feita em [21][2]. Pesquisadores de outras áreas também deram diversas contribuições para a metodologia de modelagem de agentes, uma revisão desses trabalhos é feita no livro[3].

No entanto, a ideia de se fazer modelos físicos capazes de descrever o comportamento social humano é bem antiga. Já em meados do século XIX o astrônomo Quetelet cunhou o termo *Physique Sociale* para designar seus próprios esforços no desenvolvimento de modelos matemáticos capazes de descrever diferentes aspectos sociais[4]. Uma das contribuições mais importantes dadas por ele foi a aplicação de conceitos e métodos de probabilidade e estatística, que na sua época eram usados principalmente na análise de dados astronômicos, em dados coletados da sociedade (por exemplo taxa de mortalidade). Além disso, é notável que os conceitos e os métodos matemáticos da probabilidade e estatística foram desenvolvidos por diversos matemáticos (como Pascal, Fermat, Laplace, entre outros grandes) tendo como laboratório o estudo dos jogos de azar. Em seu trabalho, Quetelet introduziu o conceito do **homem médio** (*l'homme moyen*), que seria a

[2] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646, May 2009

[3] J. Epstein. *Generative social science: Studies in agent-based computational modeling*. Princeton University Press, Princeton, 2006; and J. M. Epstein and R. Axtell. *Growing artificial societies : social science from the bottom up*. Brookings Institution Press., Washington DC, 1996

[4] J. Q. Stewart. The Development of Social Physics. *American Journal of Physics*, 18(5):239, 1950

representação matemática do homem a partir de variáveis que seguem a distribuição gaussiana. A introdução, dada por Quetelet, de conceitos estatísticos no contexto social chegou ao conhecimento de Maxwell e lhe serviu de inspiração na construção das fundações da Mecânica Estatística [5].

[5] S. Fienberg. A brief history of statistics in three and one-half chapters: a review essay. *Statistical Science*, 7(2):208–225, 1992

Atualmente, técnicas modernas empregadas em neurociência, como imagens funcionais de ressonância magnética (**fMRI**), permitem avaliar as respostas do cérebro de indivíduos quando esses são submetidos a variados tipos de estímulos e interações. Além disso, no âmbito da psicologia e sociologia, a quantidade crescente de pesquisas envolvendo dados numéricos, como por exemplo pesquisas de opinião, fornecem uma oportunidade para modelarmos alguns aspectos da sociedade.

Nessa tese, iremos modelar o comportamento do aprendizado moral de pessoas, através de um modelo de agentes Bayesianos adaptativos interagentes, baseado no modelo proposto em 2011 por Caticha e Vicente[24][6]. A nossa modelagem é fortemente baseada em evidências experimentais. Usaremos o arcabouço teórico e evidências empíricas fornecidas pela Teoria dos Fundamentos Morais, que foi proposta e desenvolvida pelo psicólogo social Jonathan Haidt juntamente com seus colaboradores em uma série de trabalhos. Essa teoria propõe que seres humanos fazem julgamentos morais de forma predominantemente intuitiva e usando um conjunto de pelo menos cinco intuições, dimensões ou fundamentos morais, sendo esses: *justiça / trapaça*, *cuidado / violência*; *lealdade / traição*; *pureza / degradação*; *autoridade / subversão*. Além disso, a aderência a esses fundamentos depende, em média, da ideologia política do indivíduo.

[6] N. Caticha and R. Vicente. Agent-Based Social Psychology: From Neurocognitive Processes To Social Data. *Advances in Complex Systems*, 14(05):711, 2011

A organização do texto principal se dá na seguinte forma: primeiramente fazemos uma revisão da literatura de ciências sociais de diversas teorias e resultados experimentais relacionados com o aprendizado moral; em seguida faremos a descrição do nosso modelo. Por fim, apresentamos e comentamos alguns resultados de simulação que estão diretamente relacionados com evidências experimentais.

⌈2⌉ Ingredientes empíricos

> It is our needs that interpret the world; our drives and their For and Against. Every drive is a kind of lust to rule; each one has its perspective that it would like to compel all the other drives to accept as a norm.
>
> — Friedrich Nietzsche,1844-1900

Por que pessoas aparentemente bem intencionadas e de culturas relativamente similares podem discordar de forma profunda sobre temas importantes para a sociedade?

A missão desse capítulo é descrever para o leitor um conjunto de teorias e experimentos que estão relacionados com o aprendizado de pessoas e que de certa forma podem ser capturados por nossa modelagem matemática. Nós vamos discutir sobre o que **"é"** a moral e como ela está relacionada com a ideologia política. Discutiremos alguns fatores determinantes para a moralidade e ideologia dos indivíduos, sendo esses fatores de natureza diversa; variando desde de possíveis influências genéticas; diversidade de informação moral que um indivíduo recebe e a maneira que se lida com novas informações.

[2.1] Cognição Moral

A conceitualização da moralidade humana é um problema no qual filósofos se debruçam há milênios. Em geral, os filósofos morais têm a preocupação de definir o que é uma atitude moral, ou como os agentes devem tomar atitudes morais enquanto o psicólogo ou o cientista moral tem a preocupação de classificar e entender diferentes tipos de comportamento moral [50, 20].

No livro *The innate mind*[1], Jonathan Haidt e Craig Joseph ressaltam logo no início do texto que durante a história dificilmente os estudiosos do comportamento moral se distanciam de sua própria moral em suas pesquisas. Um exemplo foi o embate científico ocorrido nas décadas de 70 e 80 entre as correntes teóricas de Lawrance Kohlberg e Carol Giligan. O primeiro autor identificava valores morais relacionados a *justiça/ trapaça* como suficientes para definir moralidade [75]. A segunda autora identificava que a moralidade também era derivada de valores morais relacionados com *cuidado / violência* [47]. Kohlberg e Giligan deram grandes contribuições sobre os estudos de psicologia moral, entre elas, a exploração de dilemas éticos usando a ferramenta de questionários em estudos longitudinais. No entanto, outros pesquisadores ocidentais, principalmente antropólogos, identificaram que a moralidade consiste em mais valores que justiça e cuidado. Por exemplo, para Shewder [106] seria suficiente para descrever a moralidade na maior parte das culturas um conjunto de três "éticas": autonomia, comunidade e divindade.

Dentre as muitas tradições ocidentais de filosofia moral que influenciaram as pesquisas de psicologia moral moderna podemos destacar duas grandes correntes que despontaram durante o período histórico do Iluminismo: *consequencialista* e *deontológicas*[59, 20][2]. As teorias morais deontológicas, com Emmanuel Kant (1724-1804) sendo o seu mais proeminente representante, afirmam que a moral pode ser derivada a partir do pensamento lógico ou de um pensamento racional puro, sem dar ênfase à consequência das ações. Já as teorias morais consequencialistas, entre as quais a *utilitarista* é a mais conhecida, afirmam que

[1] J. Haidt and C. Joseph. The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In *The innate mind*, volume 3, chapter 19, pages 367–391. Oxford University Press, New York, 2007

[2] J. Haidt and C. Joseph. The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In *The innate mind*, volume 3, chapter 19, pages 367–391. Oxford University Press, New York, 2007; and W. Casebeer. Moral cognition and its neural constituents. *Nature Reviews Neuroscience*, 4(October):1–6, 2003

os atos morais devem ser medidos de acordo com as consequências que eles geram na sociedade, sendo os grandes representantes dessa corrente Jonh Stuart Mill (1808-1873) e Jeremy Bentham (1748-1832). Em contra partida, outra vertente filosófica ocidental, conhecida como *teoria das virtudes morais*, que tem origem na Grécia antiga, da qual filósofos como Platão e Aristóteles são grandes representantes, afirma que o comportamento moral deve ter o intuito de cultivar virtudes e evitar vícios.

De acordo com William D. Casebeer[20], seriam necessários diferentes esforços cognitivos, que podem ser medidos experimentalmente, para executar atitudes morais em cada uma das desses três filosofias. O uso das éticas formalistas e consequencialistas demandariam que as decisões morais fossem feitas por áreas de cognição superiores como córtex pré-frontal e algumas partes destinadas a cognição sensorial. Já a ética da virtude necessitaria da coordenação dessas partes com outras associadas com o processamento de emoções. Em geral, a inferência sobre a ativação cerebral e decisão moral é feita através de estudos de **fMRI**[3] onde a atividade cerebral do indivíduo é monitorada enquanto ele deve se concentrar na resolução de dilemas éticos. Para uma revisão mais recente sobre essa metodologia indicamos a referência [25][4].

Um crescente corpo de experimentos [52, 25] evidenciam que a cognição moral necessita da coordenação tanto de regiões cerebrais relacionadas ao processamento de emoção quanto de regiões de cognição mais avançadas que estão ligadas ao planejamento de ações e raciocínio lógico, corroborando uma visão mais próxima da éticas da virtudes sobre a cognição moral [20].

Da mesma maneira que foi considerado em[24], mais importante para o nosso trabalho é o fato de que violações morais causam uma grande reação negativa associadas a ativação de áreas relacionadas a cognição emocional e social. Isso evidencia que uma parte importante do processamento moral é de origem intuitiva e automática[5], sendo que as regras de conduta moral mais intricadas seguem de uma racionalização da conduta moral que é feita a posteriori. Para

[3] Imagens de ressonância magnética funcional. Acrônimo de: *functional magnetic resonance imaging*

[4] J. F. Christensen and a. Gomila. Moral dilemmas in cognitive neuroscience of moral decision-making: a principled review. *Neuroscience and biobehavioral reviews*, 36(4):1249–64, Apr. 2012

[5] J. Haidt. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108(4):814–834, 2001; J. Greene and J. Haidt. How (and where) does moral judgment work? *Trends in cognitive sciences*, 6(12):517–523, Dec. 2002; and

uma discussão detalhada sobre os componentes emocionais da intuição moral e seus substratos cerebrais sugerimos ao leitor as referências [53, 81, 123].

Não estamos afirmando, no entanto, que emoções ou intuições são os únicos componentes por trás da cognição moral, pois como é sugerido experimentalmente, respostas negativas automáticas podem ser sobrepujadas por respostas mais utilitaristas, que são feitas recrutando-se áreas do córtex pré-frontal, e ocorrem principalmente quando pessoas devem julgar difíceis dilemas éticos pessoais com importantes consequências sociais[6]. De maneira análoga, um experimento conduzido recentemente [7] mostra que tomar decisões rapidamente privilegia cooperação, enquanto o contrario, faz com que o indivíduos tomem decisões "racionais"que favorecem a si mesmos.

Finalmente, consideraremos uma perspectiva mais descritiva do comportamento moral e que é bem representada pela definição de sistema moral dada por Jonathan Haidt[56, 60][8].

O sistema moral é um conjunto de valores, virtudes, normas, práticas, identidades, instituições, tecnologias e mecanismos psicológicos evoluídos que trabalham para agregar, suprimir ou regular o auto interesse e fazendo que sociedades operantes sejam viáveis [9].

⌈2.1.1⌉ Teoria dos Fundamentos Morais

A Teoria dos Fundamentos Morais foi inicialmente proposta pelos psicólogos sociais Jonathan Haidt e Craig Josef em [58][10]. Nesse trabalho os autores procuraram na literatura fundamentos ou dimensões morais que poderiam ser inatos a todos os seres humanos. Usando uma abordagem metaempírica os autores contabilizaram a frequência com que palavras com conteúdo moral [11] apareceram em cinco trabalhos acerca do comportamento humano. Dois trabalhos que se propunham descrever o que é universal nos humanos [16, 38], dois que analisavam o que é culturalmente variável[99, 106] e um trabalho sobre teorias evolutivas da "moral" encontrada em primatas além do *Homo Sapiens*[30]. As palavras selecionadas foram agrupadas em cinco di-

[6] D. a. Pizarro and P. Bloom. The intelligence of the moral intuitions: A comment on Haidt (2001). *Psychological Review*, 110(1):193–196, 2003; and M. Koenigs, L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser, and A. Damasio. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, 446(7138):908–11, Apr. 2007

[7] D. G. Rand, J. D. Greene, and M. a. Nowak. Spontaneous giving and calculated greed. *Nature*, 489(7416):427–430, Sept. 2012

[8] J. Haidt. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Pantheon Books, New York, 2012; and J. Haidt and S. Kesebir. Morality. In *Handbook of Social Psychology*, chapter 22, pages 797–832. Wiley, 2010

[9] tradução livre de:*Moral System are interlocking sets of values, virtues, norms, practices, identities, institutions, technologies and evolved psychological mechanisms that work to gather to suppress or regulate self-interest and make operative societies possible.*

[10] J. Haidt and C. Joseph. Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, (Special issue on human nature):55–66, 2004

[11] A lista de palavras com conteúdo moral selecionadas pelos autores pode ser encontrada na referência[48]

mensões ou fundamentos morais que são justificáveis como produtos de diferentes desafios evolutivos [12].

Justiça / Trapaça: Surge como produto do desafio evolutivo de aproveitar recompensas devido a cooperação sem que se seja explorado. Torna as pessoas sensíveis a indicativos que outra pessoa possa ser um bom (ou mau) parceiro de colaboração e altruísmo recíproco.

Cuidado / Violência: Surge como produto do desafio evolutivo de cuidar de crianças vulneráveis. Isso torna as pessoas sensíveis a sinais de sofrimento e necessidade, fazendo com que desprezem crueldade e queiram cuidar de quem está sofrendo.

Lealdade (a grupos)/ Traição: Surge como produto do desafio evolutivo de se manter coalizões. Torna as pessoas sensíveis a sinais de que outros indivíduos são ou não bons membros para o grupo. Faz com que queiram recompensar esse tipo de indivíduo e faz com que as pessoas queiram machucar, ostracizar, ou até mesmo matar aqueles que consideram traidores.

Respeito à autoridade / Subversão: Surge como resposta aos desafios evolutivos que forjam relações sociais hierárquicas. Torna as pessoas sensíveis a sinais de posição ou status social, e a sinais de que pessoas estão agindo de maneira adequada ou não de acordo com suas posições.

Santidade (ou Pureza)/ Degradação: Surge como resposta ao desafio evolutivo do dilema do onívoro [13], e também surge como resposta ao desafio mais geral de viver sob o risco de contrair doenças devido a parasitas. Essa dimensão leva em conta o comportamento do sistema imunológico fazendo com que os indivíduos se preocupem com diversos objetos simbólicos e ameaças. Faz com que pessoas possam investir valores extremos e irracionais à objetos.

[12] Os nomes dimensões morais são tradução livre respectivamente de:
Fairness / Cheating;
Care / Harm;
(in-group) Loyalty / Betrayal;
Authority / Subversion;
Sanctity (or Purity)/ Degradation.

[13] Experimentar um novo tipo de alimento que pode ser mais nutritivo (ou não) e correr o risco de ser alvo de predação ou de intoxicação alimentar, ou permanecer na mesma dieta

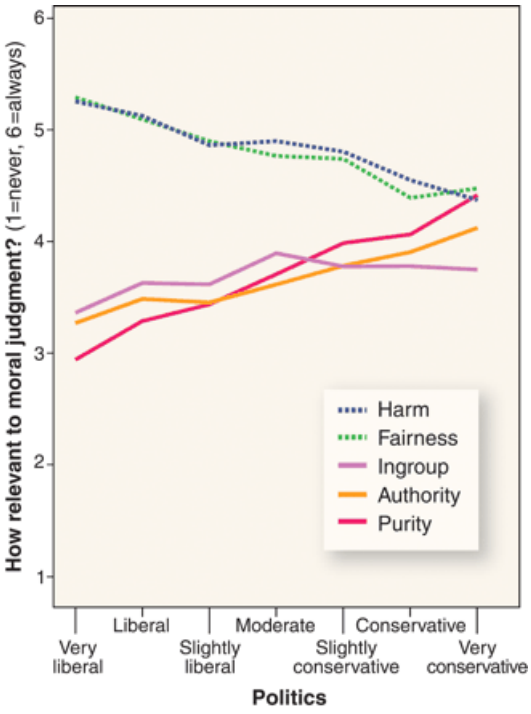


Figura 2.1: Figura retirada de [55]: Fundamentos morais de conservadores e de liberais. No eixo das coordenadas é apresentado a relevância do fundamento morais um julgamentos. No eixo das abscissas é apresentado a filiação política dos respondentes. Percebe-se que liberais julgam como mais importantes os fundamentos de Violência e Justiça, enquanto conservadores julgam com mesma importância os outros fundamentos morais, lealdade a grupo, respeito a autoridade e pureza. A gráfico é obtido a partir das repostas de cidadãos americanos do questionário hospedado em [2]. Cidadãos de outras nacionalidades apresentam resultados semelhantes.

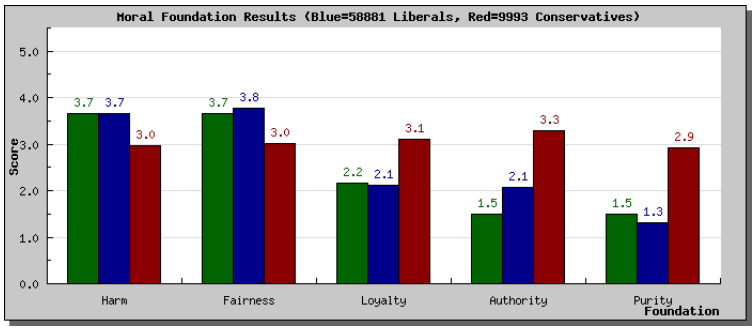
Um dos pontos mais importantes da Teoria dos Fundamentos Morais é a constatação experimental de que em média as pessoas utilizam diferentes conjuntos de fundamentos morais dependendo de sua ideologia política [55, 57][14]. Como observa-se na figura 2.1, pessoas que denominam sua filiação política como liberal tendem a fazer julgamentos morais levando em conta principalmente as dimensões justiça e cuidado; no entanto, conservadores também levam em consideração com a mesma importância as outras três dimensões de moralidade, lealdade, respeito à autoridade, santidade.

Usando um questionário, que está disponível online[2][15], é possível estimar (numa escala de 1 a 6) a importância que um indivíduo dá a cada um dos fundamentos morais. Ao responder o questionário o indivíduo primeiramente declara sua afiliação política numa escala de 1 a 7, sendo 1 muito liberal e 7 muito conservador. Na figura 2.2 vemos um exemplo de matriz moral de um indivíduo (verde) comparada com a matriz moral média de indivíduos conservadores (vermelho) e liberais (azul). O conjunto de fundamentos morais de um indivíduo é chamado pelos proponentes da Teoria dos Fundamentos Morais de

[14] J. Haidt. The new synthesis in moral psychology. *Science (New York, N.Y.)*, 316(5827):998–1002, May 2007; and J. Haidt and J. Graham. Planet of the Durkheimians, where community, authority, and sacredness are foundations of morality. In J. T. Jost, A. C. Kay, and H. Thorisdottir, editors, *Social and psychological bases of ideology*, volume Social and, chapter 15, pages 371–401. Oxford University Press, 2009

[15] Morality Quiz/Test your Morals, Values & Ethics - Your Morals.Org

matriz moral [16].



[16] O termo matriz moral é usado no sentido de fundação ou alicerce moral

Figura 2.2: Exemplo de medida da matriz moral de um indivíduo (verde), comparada com a matriz moral media de indivíduos liberais (azul) e conservadores (vermelho). A ordem de apresentação dos fundamentos da esquerda para direita é violência, justiça, lealdade, autoridade e pureza.

Tendo em vista que grande parte dos filósofos e cientistas sociais se identificam com ideologias liberais fica claro o motivo pelo qual eles definem moralidade como um conjunto de regras para o convívio social que está baseado nas dimensões justiça e cuidado.

Para tentarmos responder por que pessoas com ideologias políticas distintas usam de forma diferente as intuições morais, iremos discutir nas próximas sessões os mecanismos de aprendizado das pessoas e os estudos que investigam porque as pessoas optam por diferentes ideologias políticas.

[2.1.2] O que está faltando?

Uma pergunta evidente é se existem outras dimensões morais relevantes. A resposta para isso é provavelmente sim. Recentemente, Jonathan Haidt e seus colaboradores incluíram a dimensão moral

liberdade/opressão: que surge como resposta ao desafio evolutivo de se viver em grupos com indivíduos, que se tiverem a chance irão dominar, intimidar e constranger os outros[56].

Ao incorporar essa nova dimensão, de acordo com os proponentes da Teoria dos Fundamento Morais, é possível entender com mais profundidade outras ideologias políticas que podem estar de fora do espectro liberal/conservador. Um exemplo é a ideologia libertária, que tem liberdade/opressão como a principal dimensão moral [17].

Em nosso trabalho, levamos em consideração somente as cinco primeiras dimensões morais e o espectro político liberal/conservador

[17] R. Iyer, S. Koleva, J. Graham, P. Ditto, and J. Haidt. Understanding libertarian morality: the psychological dispositions of self-identified libertarians. *PloS one*, 7(8):e42366, Jan. 2012; and J. Haidt. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Pantheon Books, New York, 2012

para que possamos fazer a comparação entre os resultados de nossa modelagem com os dados experimentais que temos em mão.

⌈2.2⌉ Mecanismos neurológicos do aprendizado por reforço

É inegável que os seres humanos aprendem através das consequências de seus atos [64]. O aprendizado por reforço pode ser pensado como o processo de aprendizado por tentativa e erro[103]. Ou seja, esse tipo de aprendizado é caracterizado pelo senso comum de que se uma ação gera ou é seguida de uma sensação satisfatória ela tem uma maior probabilidade de ocorrer novamente, caso contrário, se a ação é seguida de um resultado negativo a probabilidade é menor[64, 11][18].

De fato, erros são importantes fontes de informação e regulação de processos cognitivos, no entanto, os mecanismos pelos quais as pessoas detectam e corrigem os seus erros não são totalmente claros, sendo alvos de estudos de grande interesse da comunidade científica[19]. Em particular, estudos de potenciais cerebrais relacionados a eventos [20] têm revelado as respostas neurais que ocorrem após erros. Esses sinais são denominados pela literatura como negatividade relacionadas ao erro [21]. A área cerebral que tem a maior probabilidade de gerar esse sinal é o córtex cingulado anterior, que também é uma área associada à monitoração de competição ou conflito [124]. A ativação dessa área também é confirmada através de imagens de **fMRI** [124]. Uma descrição mais detalhada desse tipo de experimentos será feita na seção 2.3.1

Além disso, existe uma forte relação do aprendizado por reforço e a ativação neurônios dopaminérgicos localizados nos núcleos basais (ou gânglios basais)[64, 96]. De fato, a dopamina é central em processos relacionados a recompensas, apesar do seu exato papel ainda ser controverso. Uma das hipóteses mais aceitas para o seu papel é que durante o aprendizado ela é usada como um antecipador de recompensas [39].

Alguns experimentos [73, 19] mostram que o córtex anterior cin-

[18] C. B. Holroyd and M. G. Coles. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4):679–709, 2002; and A. G. Barto. Adaptive Critics and the Basal Ganglia Adaptive Critics and the Basal Ganglia. *Models of information processing in the basal ganglia*, page 215, 1995

[19] N. Yeung, M. M. Botvinick, and J. D. Cohen. The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological Review*, 111(4):931–959, 2004

[20] tradução livre de: *Event-related brain potential*

[21] tradução livre de: *error-related negativity*

gulado é ativado quando indivíduos têm opiniões conflitantes com a norma imposta por um grupo. Isto indica que os mecanismos neuronais de aprendizado por reforço corroboram a teoria de influência social moderna que será discutida com maiores detalhes na sessão 2.4.1. É importante salientar que o córtex anterior cingulado também é ativado durante a dor física [110, 33]. Estudos recentes mostram que a dor física compartilha uma grande parte das representações somatossensorias com a percepção de forte exclusão social[76]. Com isso, é possível concluir que a exclusão social gera uma sensação correlata à dor física.

⌈2.3⌉ Ideologia Política

Os motivos dos diferentes tipos de ideologia política são fonte de intenso debate entre os cientistas sociais e principalmente entre os cientistas políticos. Além disso, mais do que explicar a diversidade de ideologias políticas, a explicação do conservadorismo soa como um enigma na aérea de ciências políticas. Em 2003, um influente trabalho de John Jost e colaboradores [70] introduziu a teoria na qual o conservadorismo político é *motivado por cognição social*.

O termo *motivado por cognição social* foi introduzido para fazer referência a que o conservadorismo, como qualquer sistema de crenças, deve ser derivado de algumas necessidades psicológicas, que variam de acordo com o ambiente (cultural, social, genético) imposto ao indivíduo. No caso do conservadorismo, as necessidades psicológicas devem estar ligadas com a manutenção do *status quo*. Neste trabalho, os autores apresentam uma metanálise de 12 pesquisas com um total 22 mil com participantes de 12 países. Eles propuseram uma teoria de psicologia abrangente incorporando várias outras teorias sobre os motivos do conservadorismo. Entre os motivos das teorias selecionadas estão motivos de personalidade (autoritárias, dogmáticas, intolerantes a ambiguidade), necessidade epistêmicas e existenciais (por fechamento, por controle de foco, por controle de medo) e racionalização ideológica (dominância social e sistemas de justificativa).

Em nosso trabalho estamos interessados em algumas características comportamentais que se diferenciam com a ideologia política e que estão relacionadas a diferentes estratégias cognitivas para se lidar com conflitos e novas informações. Por exemplo, um resultado bem conhecido nas ciências sociais relaciona ideologia política com diferentes componentes dos 5 grandes traços de personalidade [22]. Enquanto pessoas liberais apresentam traços maiores de abertura à experiência, conservadores estão positivamente correlacionados com traços de escrupulosidade[23].

Em um trabalho interessante feito em 2009 [24] Shook e Fazio examinaram a diferença do perfil exploratório entre liberais e conservadores em uma tarefa de aprendizado probabilístico. Nesse experimento, foi apresentado para os participantes, em uma tela de computador, um conjunto de imagens de sementes que variavam de aparência e tamanho. O objetivo do participante era descobrir quais os formatos de grãos são vendidos com maior lucro. Verificou-se que pessoas liberais aderiam à estratégias mais arriscadas e exploratórias, mas que conferiam uma vantagem no fim do experimento, já os conservadores aderiam a estratégias mais prudentes e menos informativas mas que lhes garantiam uma vantagem no início do experimento. Como veremos em seguida, essas características foram observadas em estudos feitos tanto no âmbito neurocientífico quanto no genético.

[22] Os 5 grandes traços de personalidades são: Neuroticismo, extroversão, sociabilidade, escrupulosidade, abertura para experiências. Eles foram introduzidos na literatura em 1961 por Tupes e Christal e pode ser encontrado em [116]

[23] A. S. Gerber, G. a. Huber, D. Doherty, C. M. Dowling, and S. E. Ha. Personality and Political Attitudes: Relationships across Issue Domains and Political Contexts. *American Political Science Review*, 104(01):111, Mar. 2010

[24] N. J. Shook and R. H. Fazio. Political ideology, exploration of novel stimuli, and attitude formation. *Journal of Experimental Social Psychology*, 45(4):995–998, July 2009

[2.3.1] Ideologia Política, Estilos Cognitivos e Neurociência

Ferramentas usadas em neurociência começaram a ser aplicadas recentemente nos estudos sobre comportamento político [69][25]. Um exemplo importante para o nosso modelo é o experimento de monitoração de conflitos e resposta à novidade feito por Amodio e colaboradores em 2007 [7][26] com o intuito de medir a diferença das estratégias cognitivas entre liberais e conservadores. Para tanto, foi executado o experimento conhecido como *Go/NoGo*. Nesse tipo de experimento os participantes devem cumprir uma tarefa quando são submetidos a um

[25] J. T. Jost and D. M. Amodio. Political ideology as motivated social cognition: Behavioral and neuroscientific evidence. *Motivation and Emotion*, 36(1):55–64, Nov. 2011

[26] D. M. Amodio, J. T. Jost, S. L. Master, and C. M. Yee. Neurocognitive correlates of liberalism and conservatism. *Nature neuroscience*, 10(10):1246–7, Oct. 2007

estímulo *Go* frequente e não cumprir a tarefa quando submetido a um estímulo *No Go* pouco frequente [27].

O estímulo *Go* é repetido de forma que o participante se acostume a ele, isso que faz com que o sinal *NoGo*, que é menos frequente, cause uma sensação de surpresa no participante, além de fazer com que ele tenha de controlar o impulso de executar a ação. O termo monitoração de conflito designa o mecanismo de detecção de quando uma resposta a um estímulo habitual entra em conflito com um estímulo imediato. Como foi discutido na sessão 2.2, a resposta cerebral a esses estímulos são comumente relacionadas ao córtex anterior cingulado [124].

Na figura 2.3(a) é mostrado a relação entre filiação política e o índice de monitoração de conflito (**ERN**) [28]. Esse índice consiste basicamente da diferença entre as amplitude máximas dos potenciais de respostas dos estímulo *Go* e *NoGo*. Com isso, vemos que o liberalismo político está positivamente correlacionado com a diferença de respostas entre os estímulos. Na figura 2.3(b) estão exemplos de curvas com os a diferença dos potenciais de resposta entre os estímulo *Go* e *NoGo* em função do tempo para conservadores e liberais. Já na figura 2.3(c) é apresentado a localização da fonte do sinal do potencial de resposta como sendo o córtex cingulado anterior. Esses resultados experimentais foram replicados de forma independente em [122].

[27] Exemplos de estímulos *Go/NoGo* podem ser a visualização de símbolos geométricos simples como quadrados, triângulos, etc, projetadas em um monitor, e um exemplo de tarefa é pressionar um botão de mouse usando um dedo

[28] acrônimo de: *Error related negativity*.

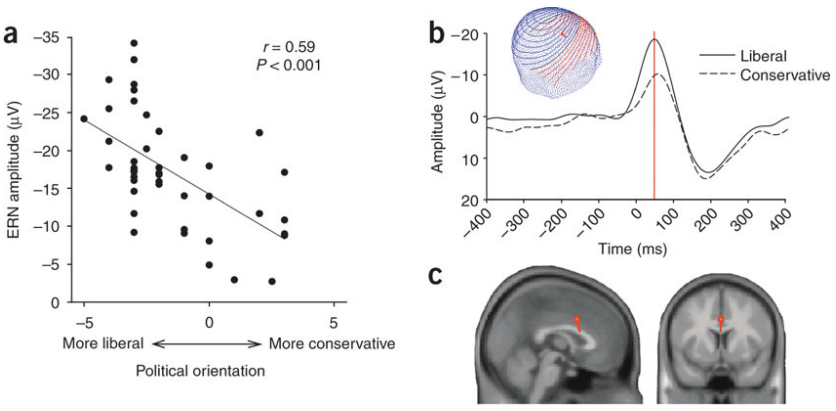


Figura 2.3: Figura retirada de [7]. Relação entre filiação política e índice de monitoração de conflito. (a) Diferenças de amplitude dos sinais (**ERN**) em função da orientação política, vemos que o liberalismo político está positivamente correlacionado com a diferença das respostas entre os estímulos *Go* e *NoGo*. (b) Exemplos de curvas com a diferença entre potenciais de resposta entre os estímulo *Go/NoGo* em função do tempo para conservadores e liberais. (c) Localização da fonte do potencial de resposta como sendo o córtex cingulado anterior.

Com isso, tem-se que indivíduos liberais tem uma maior ativação cerebral com a ocorrência de eventos inesperados quando comparados

com a ativação cerebral de eventos que são previsíveis. De maneira análoga, percebemos que conservadores tem uma maior ativação cerebral para eventos previsíveis em relação aos imprevistos quando comparado com liberais. De acordo com Jost e Amódio[69], esse resultado juntamente com outras evidências científicas, sugere que o liberalismo é associado com uma forte motivação de procura de novas informações e integração de informações conflitantes para que o indivíduo consiga entender a realidade.

⌈2.3.2⌉ Ideologia política, genética e dopamina: Primeiras evidências

Ainda é comum nas pesquisas dos cientistas sociais o paradigma de que a ideologia política é decorrente somente do contexto social. No entanto, está crescendo o número de trabalhos que relacionam a filiação política de indivíduos não somente com o meio social no qual ele está inserido mas também com outras componentes de origem biológica [41, 29, 101, 61].

Em 1984, em um estudo[29] pioneiro, o geneticista Nicolas Martin juntamente com seus colaboradores sugeriu que genes podem influenciar as atitudes de indivíduos em relação a tópicos como aborto, imigração, pena de morte, pacifismo entre outros. Eles usaram nesse trabalho a técnica clássica de estudo com gêmeos para inferência genético / comportamental: comparando gêmeos monozigóticos com heterozigóticos. Em média, gêmeos monozigóticos compartilhavam crenças politicas mais do que os heterozigóticos. Haja visto que com grande probabilidade gêmeos crescem sobre um mesmo contexto familiar, os autores chegaram a conclusão que fatores genéticos têm um papel significante sobre a atitude política dos indivíduos[30]. Apesar de ser um assunto polêmico dentro das ciências políticas, as implicações desse trabalho passaram desapercebidas durante pelo menos 20 anos. Em 2005 os cientistas políticos Hibbing e Alford reanalisaram os dados usados por Martin incorporados a outros conjuntos de dados e constataram novamente uma grande correlação entre genética e visão

[29] N. G. Martin, L. J. Eaves, a. C. Heath, R. Jardine, L. M. Feingold, and H. J. Eysenck. Transmission of social attitudes. *Proceedings of the National Academy of Sciences of the United States of America*, 83(12):4364–8, June 1986

[30] L. Buchen. Biology and ideology: The anatomy of politics. *Nature News*, 490:466, 2012

política[31]. A partir daí, uma série de outros trabalhos começaram a incorporar informações genéticas a pesquisa de ciências políticas. Recomendamos ao leitor a discussão feita por Smith e colaboradores [32] que ilustra a trajetória entre genética e atitude política incluindo 4 níveis intermediários: biológico, cognitivo/processamento de informação, personalidades/valores e ideologia levando em conta a influência de ambiente entre eles.

[31] J. Alford, C. Funk, and J. Hibbing. Are political orientations genetically transmitted. *American Political Science Review*, 99(2):153–167, 2005

[32] K. B. Smith, D. R. Oxley, M. V. Hibbing, J. R. Alford, and J. R. Hibbing. Linking Genetics and Political Attitudes: Reconceptualizing Political Ideology. *Political Psychology*, 32(3):369–397, June 2011

⌈2.3.3⌉ Ideologia política, genética e dopamina: Evidências mais recentes

Recentemente, o estudo desenvolvido pelo cientista político James Fowler e seus colaboradores no qual, usando dados do *National Longitudinal Study of Adolescent Health*, conseguiram uma associação entre a filiação politica e o número de amigos na adolescência para pessoas que possuíam duas cópias do alelo **7R** do gene de receptor de dopamina **D4** (**DRD4-7R**)[29][33].

[33] C. T. Dawes and J. H. Fowler. Partisanship, Voting, and the Dopamine D2 Receptor Gene. *The Journal of Politics*, 71(03):1157, July 2009

A dopamina é um neurotransmissor da família das catecolaminas que possui diferentes funções no cérebro que são relacionadas as funções de seus 5 tipos de receptores,(**D1,D2,...,D5**). O receptor de dopamina **DRD4** é uma proteína sintetizada por um gene que leva o mesmo nome, esse gene existe em pelo menos 3 formas polimórficas diferentes, entre elas a forma que possui o alelo **7R** [29, 114]. Existem diversos estudos que relacionam a presença desse gene com características comportamentais de busca de novidade, impulsividade, extravagância, tendência exploratória, enquanto ausência desse gene está relacionada com características como rigidez de pensamento, lealdade. Um estudo recente, indica que diferença em perfis de exploração de indivíduos podem ser previstos em média a partir de genes dopaminérgicos [42].

O programa *National Longitudinal Study of Adolescent Health* é um estudo longitudinal feito nos Estado Unidos com o intuito de coletar dados representativos da população. O estudo é feito através de questionários respondidos tanto individualmente como através de en-

trevistas feitas nas casas do participantes. O objetivo do questionário é a coleta de dados relativos aos aspectos econômicos, sociais, familiares e educacionais, e relacioná-los com os estados de saúde e bem estar desses indivíduos.

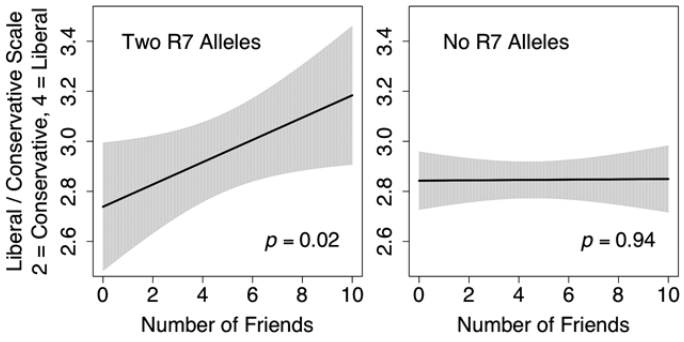


Figura 2.4: Figura retirada de [29]: A esquerda está o ajuste linear da tendência entre ideologia política e número de amigos para pessoas com duplos alelo do *R7*. A direita aponta que os autores não encontraram nenhuma relação entre números de amigos e filiação política para quem não apresentava alelos longos *R7*.

Usando os dados do *National Longitudinal Study of Adolescent Health* eles obtiveram os testes com marcadores genéticos de 2.574 indivíduos entre os quais estava o gene **DRD4**. Desse total, 33% apresentavam uma cópia do alelo **7R**, 5% duas cópias, 62% das pessoas não apresentavam cópias do alelo. A principal contribuição desse trabalho é apresentada no gráfico 2.4 onde é mostrado a correlação entre o número de amigos que o indivíduo tinha na sua adolescência e sua filiação política para pessoas com dois alelos **DRD4-7R**. Mais especificamente, como mostrado na figura esquerda de 2.4 quanto mais amigos esses indivíduos tiveram em sua adolescência maior a probabilidade desses indivíduos se tornarem liberais no início de sua vida adulta. No entanto, para indivíduos que não apresentavam nenhuma cópia do alelo, não foi encontrada nenhuma correlação entre o número de amigos (numa escala de 0 a 10) e a filiação política (numa escala de 1 a 5 sendo 1 muito conservador e 5 muito liberal). Já para pessoas que apresentavam somente um alelo, a correlação entre filiação politica e número de amigos não foi estatisticamente significante quando comparado a outras quantidades pesquisadas.

Como é salientado pelos autores, o resultado obtido nesse trabalho não é definitivo sobre o ponto de vista de uma possível relação causal entre gene e ideologia política; no entanto, ele mostra algumas pistas sobre possíveis marcadores genéticos e relações sociais que podem ser

importantes para influenciar a filiação politica do indivíduo.

Concluímos assim, que devido às evidências genéticas, de traços de personalidades, de relações sociais na adolescência, entre outras, que o liberalismo político deve estar positivamente correlacionado com a quantidade ou diversidade de informações morais a que os indivíduos foram expostos no período de maior formação cognitiva.

⌈2.4⌉ Pressão Social

Um dos grandes desafios das ciências sociais é medir a influência que os indivíduos têm entre si e que a sociedade tem sobre os indivíduos. Nesta seção falaremos um pouco sobre a teoria moderna de influência social e também discutiremos como ameaças a grupos alteram a ideologia política de indivíduos.

⌈2.4.1⌉ Influência do grupo

De acordo com [4][34] é possível definir três tipos de influência social: normativa, informacional e referente informacional (tradução livre de: *normative, informational, referent informational*). As duas primeiras (normativa, informacional) são influências interpessoais e a última (influência referente informacional) está relacionada à noção de pertencer a um grupo. A influência normativa é caracterizada quando um indivíduo expressa opiniões ou age publicamente de acordo com um padrão para evitar punição social ou obter alguma recompensa social. A influência informacional é o mecanismo de influência comum quando membros de um grupo experienciam incerteza subjetiva e a falta de evidências objetivas para se avaliar um estímulo. A influência social surge pois, a fim de diminuir a incerteza, o indivíduo engaja em comparações sociais com os outros membros do grupo. Já a influência referente informacional contabiliza o quanto a percepção de pertencimento a um grupo muda a opinião do indivíduo.

[34] D. Abrams, M. Wetherell, S. Cochrane, M. a. Hogg, and J. C. Turner. Knowing what to think by knowing who you are: self-categorization and the nature of norm formation, conformity and group polarization. *The British journal of social psychology / the British Psychological Society*, 29 (Pt 2)(May 1987):97–119, June 1990

Os experimentos clássicos de Sherif e Asch [104, 9][35] exemplificam a influência que grupos têm sobre indivíduos[24]. No experimento de Sherif, os participantes são colocados em uma sala escura e têm a tarefa

[35] M. Sherif. An experimental approach to the study of attitudes. *Sociometry*, 1(1):90–98, 1937; and S. Asch. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, 70(9), 1956

de julgar se um ponto de luz projetado na parede está se movimentando ou não, sem que eles saibam que de fato a projeção é estática. Esse tarefa é repetida diversas vezes com os participantes sozinho na sala ou em grupo. Quando em grupo, as estimativas dos participantes convergem para um padrão especifico, ou seja, os membros do grupo acabam entrando no consenso de que a projeção está parada ou se movimentando.

No experimento de Asch, os participantes têm de comparar tamanhos de listras verticais apresentadas em duas cartas. Na primeira carta existe 1 linha vertical como um comprimento de referência. Na segunda carta existem 3 linhas verticais com tamanhos distintos onde uma das linhas tem o mesmo comprimento da linha de referência, como pode ser visto na figura 2.5. O participante tem de julgar qual das linhas da segunda carta tem o mesmo tamanho da linha de referência da primeira carta.

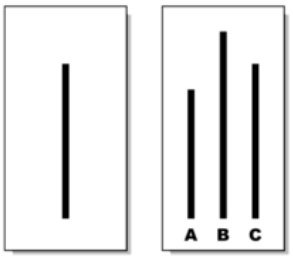


Figura 2.5: Par de cartas apresentadas nos experimento do Asch. À esquerda a linha de referência e à direita as três linhas para comparação. Figura retirada de [1].

O participante é colocado numa sala juntamente com um grupo de *confederados*, que sabem o verdadeiro objetivo do experimento. Ele deve expressar sua opinião sobre a tarefa somente depois de ouvir a opinião de todos os confederados. A influência do grupo é verificada pois os confederados escolhem entre as opções de linhas verticais uma alternativa que é claramente errada. Isso faz com que a maioria dos participantes escolham a mesma opção do grupo.

O experimento de Asch é usualmente interpretado como um exemplo de predominância da influência normativa, já que existe uma norma social clara e os participantes devem expressar uma opinião contrária à que teriam caso não tivessem de expressá-la publicamente. No entanto, essa interpretação não está fora de questionamento, pois, como argumenta [4], pode-se pensar que o objetivo do experimento não é tão claro para o participante; sendo assim, a influência predominante é do tipo informacional e ocorre através de complacência, ou seja, apesar de não concordar com o grupo, o participante se submete a ele com o intuito de diminuir sua incerteza sobre o estímulo.

Já o experimento de Sherif é um exemplo clássico no qual a influência informacional age de forma preponderante pois o grupo cria uma

norma própria a partir da análise de um estímulo ambíguo, e, além disso, o estímulo é realizado numa sala totalmente escura, fazendo com que os indivíduos participem do experimento de forma anônima e discreta para os outros membros do grupo.

Como estudado experimentalmente em [4], os dois tipos de influências interpessoais (normativa e informacional) dependem de que os indivíduos se percebam como membros do grupo. Para isso os autores do estudo refizeram os experimentos de Sherif e Asch com a diferença de que à noção de pertencimento de grupo dos participantes era salientada. Com isso, eles sugerem que os tipos de influência social interpessoal são casos particulares de influência referente informativa quando o indivíduo tem completa noção de pertencimento ao grupo. Uma revisão mais atual sobre os processos de influência social pode ser encontrada em [26].

⌈2.4.2⌉ Ameaças e conservadorismo

De acordo com [46][36], a Teoria de controle de terror [37] [49] providencia umas das mais proeminentes explicações para intolerância e viés contra diferenças pessoais. Por essa teoria, o instinto de auto preservação aliado com a consciência da mortalidade cria no indivíduo um potencial terror paralisante [8, 109]. Para controlar esse terror, pessoas têm a necessidade de ter fé em visões de mundo estáveis, e se sentirem membros com valor dentro de um universo que têm algum sentido. Essa visão de mundo é culturalmente defendida através de crenças sobre a realidade. Com isso, as pessoas são fortemente motivadas em manter a fé em suas visões de mundo e agirem de acordo.

Dentro da teoria de manutenção de terror a hipótese da saliência da mortalidade diz que ao relembrar um indivíduo de sua mortalidade ele tende a aumentar sua necessidade pelas estruturas de sua visão de mundo. Diversos estudos corroboram essa hipótese demostrando que estímulos relacionados com morte resultam na defesa da visão de mundo principalmente através de reações negativas em relação a pessoas diferentes e reações positivas em relação a pessoas similares [97, 18]. Portanto, uma possível implicação da teoria de

[36] A. E. Giannakakis and I. Fritsche. Social identities, group norms, and threat: on the malleability of ingroup bias. *Personality & social psychology bulletin*, 37(1):82–93, Jan. 2011

[37] tradução livre de:*Terror management theory*

controle de terror no âmbito de ideologia política é que ameaças fariam com que liberais se tornassem mais liberais e conservadores mais conservadores[83][38].

No entanto, o modelo de cognição social motivada [70] prevê que ameaças fazem com que tanto liberais quanto conservadores tenham uma tendência ao conservadorismo. De fato, também existe uma extensa literatura [39] que corrobora a ideia de que situações de ameaças aumentam as tendências conservadoras dos indivíduos independentemente de sua ideologia política.

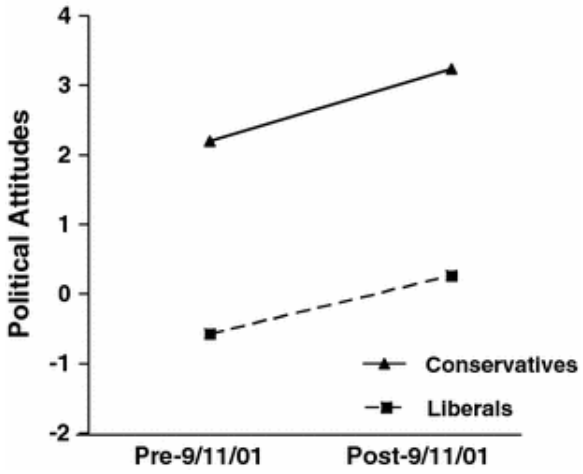


Figura 2.6: Figura retirada de [83]. Índice de atitude política média medida antes e depois dos ataque terrorista de 11/09/2001. Quanto maior o índice de atitude política maior a ideologia conservadora.

Por exemplo, alguns estudos apontam que ameaças à sociedade como os atos terroristas de 11 de setembro tornam as pessoas mais conservadoras[15, 83]. Na figura 2.6 apresentamos o resultado principal do artigo [83]. Neste trabalho, através de um questionário foi medido um índice de atitude politicas antes e depois do atentado de 11/09/2001 para diferentes grupos de pessoas que auto denominaram sua ideologia política.

O efeito de crescimento do conservadorismo também pode ser obtido em condições de laboratório em experimentos onde os participantes são induzidos a se sentirem ameaçados, por exemplo, fazendo os participantes pensarem sobre situações de injustiça ou salientando suas mortalidades [84][40].

[38] P. R. Nail and I. McGregor. Conservative Shift among Liberals and Conservatives Following 9/11/01. *Social Justice Research*, 22(2-3):231–240, June 2009

[39] Consultar [68, 83] e suas respectivas referências.

[40] P. R. Nail, I. McGregor, A. E. Drinkwater, G. M. Steele, and A. W. Thompson. Threat causes liberals to think like conservatives. *Journal of Experimental Social Psychology*, 45(4):901–907, July 2009

⌈3⌉ Modelo

> Every attempt to employ mathematical methods in the study of chemical questions must be considered profoundly irrational and contrary to the spirit of chemistry... if mathematical analysis should ever hold a prominent place in chemistry – an aberration which is happily almost impossible – it would occasion a rapid and widespread degeneration of that science.
>
> — Auguste Comte, 1798-1857

Este trabalho é uma extensão do modelo que foi inicialmente apresentado por Caticha e Vicente [24][1] para o estudo da dinâmica moral da sociedade. O trabalho de Caticha e Vicente é descrito de forma mais detalhada e dentro do contexto mais geral de dinâmica de opinião no apêndice C.1, onde também apresentamos algumas modificações e cenários de simulação trabalhados durante o período de doutoramento. As diferênças entre as duas modelagens serão discutidas ao longo do texto.

[1] N. Caticha and R. Vicente. Agent-Based Social Psychology: From Neurocognitive Processes To Social Data. *Advances in Complex Systems*, 14(05):711, 2011

Neste seção, apresentaremos a extensão do modelo de [24] para agentes Bayesianos e a análise de um conjunto de dados fornecidos por Jonathan Haidt *et al.*[2] aparecerá no capítulo 4. Alguns anos mais tarde, e após este trabalho já ter sido concluído, tivemos acesso a um conjunto de dados muito mais extenso onde há dados referentes ao mesmo país de respondentes mas também a um conjunto muito maior de nações. Usando o modelo original[24] fizemos uma análise multi cultural apresentada no capítulo 5.

No modelo, a **matriz moral** de um agente i é um vetor $\boldsymbol{\omega}_i = (\omega_{i1}, \ldots, \omega_{i5})$ de 5 dimensões, onde cada componente ω_{ia} pode ser interpretada

como uma dimensão moral. Assumimos, por simplicidade, que as pessoas são igualmente morais, ou seja, a diferença entre morais de indivíduos é expressa somente através da direção no espaço moral. Para isso, faremos que os vetores morais de nossos agentes sejam de modulo $|\boldsymbol{\omega}_i| = 1$.

Uma das hipóteses mais fortes do nosso trabalho é que a descrição matemática do aprendizado moral pode ser dividida em duas fases, sendo que cada uma das fases necessita de modelagens distintas. A **fase 1** mimetiza o aprendizado moral de pessoas no período da infância até a adolescência. Durante essa fase o julgamento de assuntos com conteúdo moral muda a estratégia cognitiva do agente. Já na **fase 2**, representamos o aprendizado moral de indivíduos adultos, ou de indivíduos que já tem uma estratégia cognitiva estabilizada. Nessa fase, estamos interessados nos estados estacionários e nos mecanismos de controle da dinâmica de opiniões da sociedade.

Em ambas as fases um agente i modifica sua matriz moral ao receber informações através da interação social com outro um agente j. A interação social se dá através da discussão de assuntos com conteúdo moral, onde a informação recebida pelo agente i ao interagir com o agente j é escrita na forma $y_\mu = (\sigma_{j\mu}, \boldsymbol{x}_\mu)$. Chamamos de **assunto** os vetores $\boldsymbol{x}_\mu = (x_{\mu 1}, \ldots, x_{\mu 5})$, onde suas componentes representam o peso ou conteúdo de um fundamento moral de um assunto discutido entre duas pessoas numa sociedade real. O termo $\sigma_{j\mu} = \pm 1$ é a **classificação** que o agente j dá ao assunto – positiva/negativa ou favorável/contrário.

Como foi discutido no capítulo 2, esperamos que julgamentos morais sejam feitos de maneira rápida e intuitiva. Para tanto, faremos com que a classificação que o agente dá a um assunto seja feita pelo sinal do produto escalar entre os vetores da matriz moral do agente e do conteúdo moral do assunto, ou seja, $\sigma_{j\mu} = sign(h_{j\mu})$ onde $h_{j\mu} = \boldsymbol{\omega}_j \cdot \boldsymbol{x}_\mu = \sum_{a=1}^{5} \omega_{ja} x_{\mu a}$. O termo $h_{j\mu}$ é o peso da classificação ou **opinião** do agente em relação ao assunto discutido. Outra importante variável é $z_\mu = h_{i\mu}\sigma_{j\mu}$ que expressa a concordância que o agente i tem em relação à classificação do agente j sobre o assunto $\boldsymbol{x}_\mu$.

A troca de informação entre os agentes do modelo proposto por Caticha e Vicente[24] ocorre de maneira similar; no entanto, a diferença fundamental entre esse modelo e o apresentado no presente trabalho é que no primeiro não existe nenhum mecanismo de construção da estratégia cognitiva do agente, o que faz com que ele só tenha uma fase de aprendizado, que é similar à segunda fase de nossa modelagem.

[3.1] Fase 1: Criação de estratégia cognitiva

Durante a fase 1 o agente definirá sua estratégia cognitiva através de trocas de informações com seus parceiros sociais. Como já discutimos anteriormente, o agente interage com seus parceiros sociais e obtém informações com conteúdo moral na forma $(\boldsymbol{x}_\mu, \sigma_\mu)$. Tendo em vista que o agente recebe esse tipo de informação, iremos fazer uma descrição probabilística de seu aprendizado usando um algoritmo de aprendizado Bayesiano aproximado, já que esse respeita critérios de razoabilidade. No apêndice B, apresentamos a motivação geral para o uso de descrições probabilísticas assim como a dedução das equações do aprendizado Bayesiano.

Sobre o ponto de vista de psicologia, podemos interpretar o algoritmo Bayesiano como uma descrição motivacional do aprendizado, no qual o agente é motivado a aprender para diminuir um custo psicológico $\mathcal{E}$. Outra interpretação equivalente pode ser feita sobre a perspectiva de aprendizado por reforço, onde através de um algoritmo do tipo Hebbiano[2], o indivíduo responde de acordo com um protocolo automático interno representado por uma função F de modulação do aprendizado[3]. O formato da função de modulação do aprendizado Bayesiano é apresentado na figura 3.1. Na fase 1, as equações que descrevem o aprendizado do agente i a partir da informação dada pelo

[2] A ideia de aprendizado Hebbiano é descrita no apêndice A

[3] Podemos observar pela equação 3.1 que a função de modulação e o custo psicológicos estão relacionados através da expressão $F = -C\partial_z\mathcal{E}$.

agente j são,

$$\begin{aligned}
\overline{\omega}_i(t+1) &= \overline{\omega}_i(t) - C_i(t)\frac{\partial \mathcal{E}(t)}{\partial \overline{\omega}_i(t)}; \\
&= \overline{\omega}_i(t) - \sigma_{j\mu}\boldsymbol{x}_\mu C_i(t)\frac{\partial \mathcal{E}(t)}{\partial z_\mu}; \\
&= \overline{\omega}_i(t) + \sigma_{j\mu}\boldsymbol{x}_\mu F(t); \qquad (3.1) \\
C_i(t+1) &= C_i(t) - C_i(t)^2\frac{\partial^2 \mathcal{E}(t)}{\partial z_\mu^2}; \\
&= C_i(t) + C_i(t)\frac{\partial F(t)}{\partial z_\mu}. \qquad (3.2)
\end{aligned}$$

Observamos pela equação 3.1 que a matriz moral do agente muda na direção do assunto discutido multiplicado pela classificação do seu parceiro social[4]. Note que, no processo de aprendizado, o agente além de mudar sua matriz moral $\boldsymbol{\omega}_i$ também muda sua **estratégia cognitiva** através da função $C_i(t)$. Como foi demostrando no apêndice B, esta função decresce a medida que o aprendizado da fase 1 ocorre e está relacionada com o inverso da correlação da distribuição de probabilidade posteriori do vetor de pesos da matriz moral do agente.

[4] Note que a equação 3.1 é vetorial enquanto a equação 3.2 é escalar.

Mais especificamente, consideramos que as distribuições a posteriori de matrizes morais devem ser projetadas sobre a família de gaussianas com a matrizes de covariâncias escritas na forma $\boldsymbol{C} = \mathbf{1}C$, onde C é uma constante e $\mathbf{1}$ é uma matriz identidade. Com essa aproximação implicitamente assumimos que não existe correlação entre as diferentes dimensões morais.

Nesse ponto, definiremos uma nova grandeza que também se relaciona com o aprendizado do agente,

$$\rho_i(t) = \frac{1}{\sqrt{1 + C_i(t)^2}}, \qquad (3.3)$$

está limitada no intervalo $(0, 1)$ e é crescente com o aprendizado do agente, ou seja, quando ρ assume valores próximos de zero temos que o agente fez poucos julgamentos morais e vai se aproximando de um a medida que o mesmo ganha esse tipo de experiência.

Para representar a infância de uma pessoa, consideramos que a estrutura social relevante para o agente i é menos complexa que na se-

gunda fase. Ou seja, a interação do agente é feita com um número menor de parceiros sociais relevantes para o seu aprendizado. Além disso, assumimos que esses parceiros sociais tem matrizes morais similares, e que não mudam muito através da interação com o agente i. Com esse procedimento, garantimos que a dinâmica de aprendizado do agente se assemelhe a uma dinâmica do tipo professor aluno, que está melhor descrita no apêndice A.

⌈3.1.1⌉ Função de modulação

A função de modulação mede a importância da informação contida nos assuntos discutidos. Podemos especular que na biologia essa função pode representar o sinal de alguma estrutura cerebral relacionada a julgamentos automáticos, como amídala ou o córtex anterior cingulado, e que ficam intensos em situações de surpresa[5],

$$F(z) = \frac{(1-2\epsilon)\exp\left(-\frac{z^2}{2C^2}\right)}{\epsilon + (1-2\epsilon)\Phi\left(\frac{z}{C}\right)}. \tag{3.4}$$

[5] $\Phi(x)$ é a função acumulada de $-\infty$ até x de uma gaussiana $\mathcal{N}(0,1)$.

O termo ϵ da função de modulação é uma estimativa da taxa de erros multiplicativos[6] na comunicação entre a classificação do assunto emitida pelo parceiro social e a percebida pelo agente de que está aprendendo [72][7]. A importância que o agente dá a um assunto está relacionada com o termo $\rho(t) = 1/\sqrt{1+C(t)^2}$, que é uma medida da capacidade de predição do agente e também uma medida da quantidade de informação julgada por ele[71][8].

[6] Erro multiplicativo significa que o sinal da opinião tem alguma probabilidade de ser invertido.

[7] O. Kinouchi and N. Caticha. Learning algorithm that gives the Bayes generalization limit for perceptrons. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, 54(1):R54–R57, July 1996

[8] O. Kinouchi and N. Caticha. Optimal generalization in perceptrons. *Journal of Physics A: Mathematical and*, 25:6243–6250, 1992

O efeito da aquisição de informação pode ser percebido na figura 3.1 onde mostramos a função de modulação do aprendizado Bayesiano em duas situações e para diferentes valores do parâmetro de socialização ou experiencia moral ρ que cresce no sentido da seta $\downarrow$. Nessa figura à esquerda é considerado o caso onde não existe erro de comunicação ($\epsilon = 0$). À medida que o aprendizado ocorre ($\downarrow$) a função de modulação decresce para assuntos em que existe concordância de opinião ($z > 0$) e aumenta para assuntos em que existe discordância de opinião ($z < 0$). Sendo assim, interpretamos que assuntos que trazem mais informação, que são os assuntos discordantes, causam mais

impacto no aprendizado do agente à medida que o mesmo acumula experiência. Outro efeito que podemos notar é que para valores pequenos de ρ o agente terá uma tendência mais corroborativa pois esse dará bastante importância para classificação de assuntos que estão de acordo com sua opinião ($z > 0$) [71].

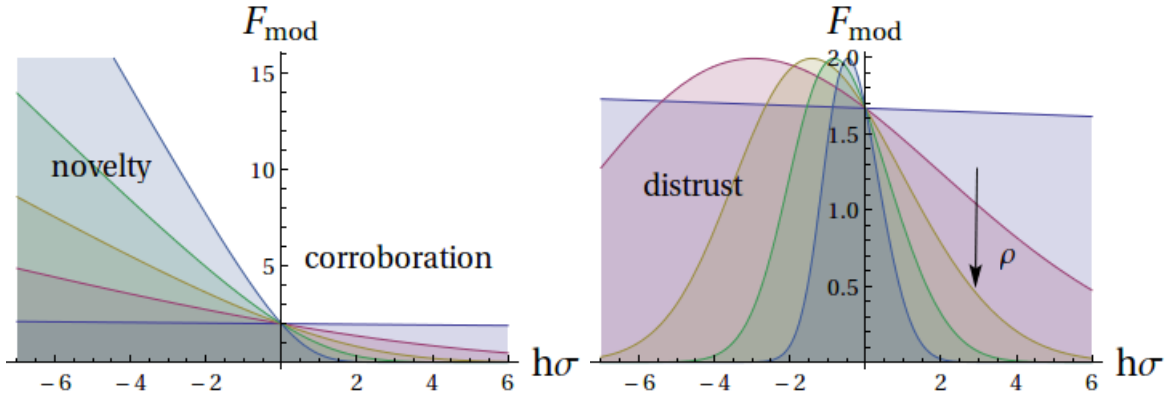


Figura 3.1: Dependência da função de modulação com a quantidade de troca de informação medido através de ρ. À direita é apresentado a função de modulação quando $\epsilon = 0$ e à esqueda quando $\epsilon = 0.2$. Quanto mais informação é processada (ρ cresce no sentido indicado pela seta $\downarrow$) maior é a amplitude da função de modulação para informações que tem novidade, $h\sigma < 0$; e menor é a amplitude para informações que não tem novidade, $h\sigma > 0$.

Já na figura à esquerda, estamos analisando o caso onde a opinião do parceiro social é invertida com probabilidade $\epsilon = 0.2$. O algoritmo Bayesiano permite incorporar essa informação, o que dá origem a um efeito muito interessante de desconfiança. Ou seja, se o agente discorda da opinião do seu parceiro social sobre um assunto e além disso está muito certo de sua própria opinião ($z \ll 0$) ele irá aprender pouco com essa nova informação. É interessante notar que o efeito de desconfiança fica mais pronunciado à medida que o agente acumula experiência, ou seja, vemos uma maior depleção na forma da função de modulação para valores negativos de z à medida que ρ cresce [72].

⌈3.2⌉ Fase 2: Aprendizado da matriz moral

Durante a primeira fase o agente aprendeu a aprender, ou seja, nesta fase ele estabeleceu a estratégia cognitiva para lidar com novas informações. Na fase 2, consideramos que o agente tem uma estratégia de aprendizado fixa, o parâmetro que mede a complexidade da socialização na primeira fase se torna independente do tempo, $\rho_i(t) = \rho_i$. A modelagem dessa transição foge do escopo desse trabalho, sendo tema de futuras pesquisas; no entanto, podemos especular que, do ponto de

vista da psicologia, ela tem alguma relação com as fases de aprendizado de Piaget[93], ou com a dificuldade da criança e adolescentes tem de se imaginar sob a perspectiva de terceiros[82, 14][9].

[9] Y. Moriguchi, T. Ohnishi, T. Mori, H. Matsuda, and G. Komaki. Changes of brain activity in the neural substrates for theory of mind during childhood and adolescence. *Psychiatry and clinical neurosciences*, 61(4):355–63, Aug. 2007; and S.-J. Blakemore. The social brain in adolescence. *Nature reviews. Neuroscience*, 9(4):267–77, Apr. 2008

Assumimos ainda que os agentes são influenciados somente por aqueles que tem o mesmo tipo de estratégia cognitivas, ou seja, $\rho_i = \rho$ para todo o agente i da sociedade. Apesar de forte, essa hipótese pode ser justificada pois indivíduos tem a tendência de considerar mais importantes as opiniões das pessoas que são parecidas consigo. Essa hipótese equivale a assumir para o nosso modelo uma espécie de influência informacional saliente [4] através do parâmetro ρ. No entanto, os efeitos dessa hipótese não são tão relevantes para os resultados do modelo pois, como mostramos no apêndice C.4, uma sociedade de agentes com diferentes estratégias cognitivas não muda de maneira drástica o perfil estatístico da opinião de um grupo de agentes que têm a mesma estratégia cognitiva.

De maneira similar à proposta em [24], simplificamos a dinâmica do modelo assumindo que o conjunto de assuntos $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_P\}$ discutidos pelos agentes podem ser condensados, ou resumidos, em um assunto médio $\mathcal{Z}$,

$$\mathcal{Z} \propto \frac{1}{P} \sum_{\boldsymbol{x}_\mu \in \mathcal{X}} \boldsymbol{x}_\mu.$$

Como o assunto médio captura a essência de todos os assuntos discutidos na sociedade, o chamamos de *Zeitgeist*[10]. Sem perda de generalidade o vetor *Zeitgeist* tem uma direção fixa e é normalizado ($\mathcal{Z} = (1,1,1,1,1)/\sqrt{5}$). A opinião de um agente i em relação ao *Zeitgeist* é definida como,

[10] Do alemão *zeit* significa tempo e *geist* significa espirito ou mente, o termo *Zeitgeist* é comumente traduzido como espirito do tempo. Para mais informações a respeito do termo olhar [40]

$$h_i = \mathcal{Z} \cdot \boldsymbol{\omega}_i.$$

Na fase 2, as conexões sociais entre os agentes são definidas através de um grafo $\mathcal{G}(\mathcal{V}, \mathcal{A})$ onde $\mathcal{V}$ é um conjunto de vértices na qual estão localizados os agentes, e $\mathcal{A}$ é o conjunto de arestas, que representa as conexões sociais entre os agentes propriamente dita.

A dinâmica do modelo na Fase 2 é similar a uma dinâmica de Metropolis [80]. Em um tempo de simulação t dois agentes vizinhos (i, j) são sorteados do conjunto de arestas $\mathcal{A}$. Um dos agentes vizinhos

é sorteado, suponha que seja o agente i, para tentar mudar sua matriz moral ω_i para uma matriz moral tentativa ω_i' através da interação social com seus vizinhos[11]. A matriz moral tentativa ω_i' é sorteada uniformemente de uma hiperesfera unitária de 5 dimensões, e sua aceitação como uma nova matriz moral pelo agente i ocorre se a energia de interação[12] entre o agente e seus vizinhos diminuir. Caso contrário, o agente aceita a matriz moral tentativa com um probabilidade igual a exponencial de menos a variação de energia vezes uma constante β positiva. Ou seja,

$$\omega_i \to \omega_i' \text{ com prob. } p = \min\{1, \exp(-\beta \Delta \mathcal{E})\}$$

Seja $\mathcal{V}_i$ a vizinhança do agente i, então escrevemos variação de energia de iteração do agente i em relação aos seus vizinhos como

$$\Delta \mathcal{E} = \sum_{k \in \mathcal{V}_i} \left[\mathcal{E}(h_i', \sigma_k) - \mathcal{E}(h_i, \sigma_k) \right]$$

Chamamos a constante β de **pressão social** ou **pressão de pares**. Esse parâmetro tem um papel fundamental na dinâmica, pois ele indica o nível de flutuação aceitável para as matrizes morais dos agentes. Para valores de β grandes, o agente terá pouca liberdade para mudar sua matriz moral; já ao contrário, para valores de β próximos de zero, o agente com grande probabilidade pode assumir uma matriz moral que aumente a energia de interação com sua vizinhança. De fato, essa interpretação faz sentido do ponto de vista social, já que é senso comum que sociedades com grande pressão social não permitem uma grande variação de comportamento. A fase 2 de nosso modelo é flexível o suficiente para assumirmos que cada agente tenha uma percepção diferente da pressão social; no entanto, evidências experimentais[13] indicam que a pressão social não é percebida com grande diferença entre indivíduos numa mesma sociedade e também não varia muito com o tamanho da comunidade que o indivíduo está inserido.

[11] Fazemos esse procedimento para o sorteio do agente pois isso garante que o número de interações sociais é proporcional ao número conexões sociais do agente

[12] A expressão da energia de iteração de um agente i com seu vizinho j é: $\mathcal{E}(h_i, \sigma_j) = \log\left(\epsilon + (1 - 2\epsilon)\Phi\left(\frac{z}{\bar{c}}\right)\right)$

[13] C. Panagopoulos. Social pressure, surveillance and community size: Evidence from field experiments on voter turnout. *Electoral Studies*, 30(2):353–357, June 2011

[3.3] Parâmetros de Ordem e transição de fase

Duas grandezas importantes para medirmos a dependência das matrizes morais dos agentes com os parâmetros do modelo, são a média m e variância v das opiniões dos agentes em relação ao *Zeitgeist*,

$$\begin{aligned} m &= \langle h_i \rangle \,, \\ v &= \left\langle (h_i - \langle h_i \rangle)^2 \right\rangle ; \end{aligned} \tag{3.5}$$

Essas grandezas são conhecidas como *parâmetros de ordem* já que as mesmas trazem informação da organização macroscópica do sistema. A dependência dessas grandezas com a pressão social β e com a medida de complexidade de socialização ρ pode ser analisada através da figura 3.2. Como já foi dito, para valores pequenos do parâmetro β existe uma grande dispersão das matrizes morais dos agentes, na figura 3.2(b) pode-se observar que à medida que a pressão social aumenta a variância das opiniões dos agente diminui.

(a) média das opiniões

(b) variância das opiniões

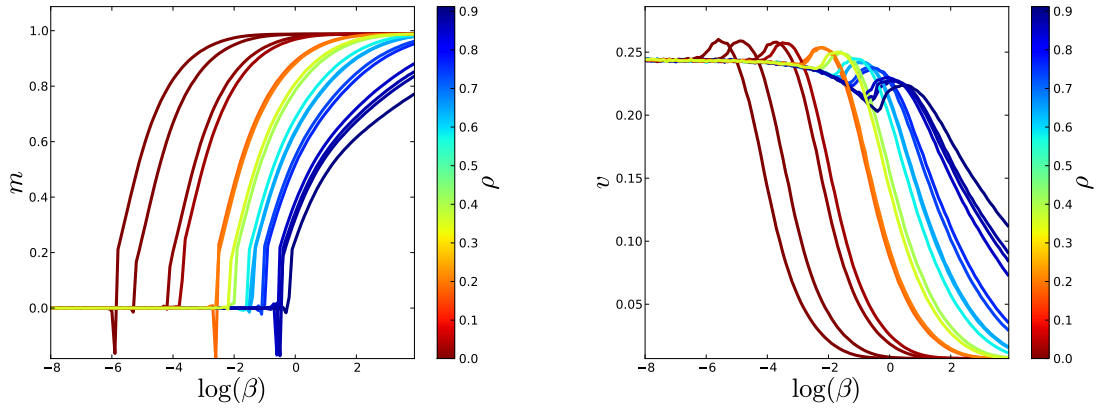


Figura 3.2: Dependência dos parâmetros de ordem m e *sigma* em função da pressão social β para vários valores de complexidade de socialização ρ. Em (a) percebemos que existe uma transição de fase de um estado desordenado $m = 0$ para um estado ordenado $m > 0$ à medida que a pressão social aumenta. Em (b) observa-se que a dispersão das opiniões na sociedade σ diminuem com o aumento da pressão social. Além disso, quanto menor o valor do parâmetro de socialização ρ menor é o valor da pressão social necessária para a transição de fase.

Um comportamento mais interessante pode ser notado através da figura 3.2(a), onde mostramos a média das opiniões dos agentes em função de β. Para valores pequenos de pressão social, a opinião média dos agentes é nula ($m = 0$); no entanto, para valores de pressão acima de um limiar $\beta_c(\rho)$ as matrizes morais dos agentes se alinham na dire-

ção do *Zeitgeit*, de forma que a opinião médias do agentes passa a ser maior que zero ($m > 0$). É importante notar que a média das opiniões dos agentes muda de maneira abrupta fazendo com que a derivada da opinião média em relação à pressão social divirja, o que caracteriza uma *transição de fase* de 2ª ordem.

Como em nosso modelo as interações entre os agentes não são simétricas ele não é um sistema Hamiltoniano, apesar disso, podemos definir o parâmetro de ordem

$$\chi = \beta \left\langle (m - \langle m \rangle)^2 \right\rangle , \tag{3.6}$$

esse parâmetro é análogo à susceptibilidade magnética do modelo de Ising, e nada mais é do que a variância da média das opiniões vezes a pressão social. A utilidade desse parâmetro em nosso modelo é que ele caracteriza a transição de fase pelo seu comportamento divergente, como pode ser notado através da figura 3.3.

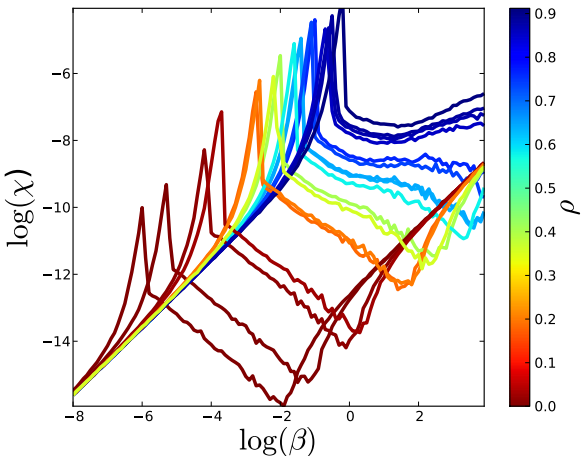


Figura 3.3: Divergência da "susceptibilidade" das opiniões dos agentes. Observa-se que essa quantidade diverge para diferentes valores de ρ para diferentes valores de β. Isso caracteriza uma transição de fase em $\beta_c(\rho)$ a ser discutida no próximo capítulo.

Por fim, conseguimos notar que a organização dos agentes em torno do *Zeitgeist* tem uma grande dependência com o parâmetro ρ, observamos que quanto maior é a socialização (ou quantidade de informação) que o agente obteve na fase 1, maior o valor da pressão social necessário para que ocorra transição de fase. Indicamos aos leitores interessados em fenômenos de transição de fase as referências[14].

[14] M. Newman and G. Barkema. *Monte Carlo methods in statistical physics*. Oxford University Press, Oxford, 1999; and H. Nishimori and G. Ortiz. *Elements of Phase Transitions and Critical Phenomena*. Oxford University Press, Oxford, 2011

[3.4] Comparação entre dados simulados e experimentais

Uma sociedade de agentes é caracterizada pelo parâmetro β que fornece a pressão social e pelo parâmetro ρ de socialização. Assim, para tempos longos da Fase 2, a sociedade de agentes apresentará uma distribuição de opiniões sobre o *Zeitgeist* $P_S(h|\rho,\beta)$ estável, apesar das opiniões individuais dos agentes não serem constantes. Usando um conjunto de dados contendo aproximadamente 15 mil questionários obtidos através de [2] podemos levantar matrizes morais empíricas normalizadas $\boldsymbol{\omega}_e$ de indivíduos separados por suas ideologias políticas, numa escala de **0** a **7** sendo **0** muito liberal e **7** muito conservador.

Inferimos o *Zeitgeist* experimental através da matriz moral média dos indivíduos mais conservadores. Usamos esse procedimento pois, de acordo com nossa teoria, esperamos que pessoas conservadoras apresentem menos dispersão de matrizes morais. Obtemos a informação das opiniões de cada respondente sobre o *Zeitgeist*, que é dada por $h_e = \boldsymbol{\omega}_e \cdot \mathcal{Z}_e$. Dessa forma, tem-se uma assinatura estatística dos dados experimentais através de histogramas ou distribuições das opiniões empíricas dado a ideologia política, $P_E(h|\text{i.p.})$.

Usando a metodologia de [15], medimos o quão próximo os histogramas de opiniões de agentes obtidos por simulação estão dos histogramas empíricos através da distância definida por

$$D(\rho,\beta;ip) = \sum_{h\in bins} \left(P_S(h|\rho,\beta) - P_E(h|ip)\right)^2 \tag{3.7}$$

que nada mais é soma da diferença quadrática das frequenciais das opiniões sobre um conjunto de bins[16]. As figuras 4.1, 4.2 e 4.3, que serão apresentadas no próximo capítulo; são obtidas identificando os histogramas experimentais mais próximos dos simulados com a restrição de que a distância $D(\rho,\beta;ip)$ seja menor que um certo valor limiar[17].

[15] N. Caticha and R. Vicente. Agent-Based Social Psychology: From Neurocognitive Processes To Social Data. *Advances in Complex Systems*, 14(05):711, 2011

[16] o termo bin se refere a um termo de um conjunto de divisões ou espaçamentos sobre um intervalo

[17] nas figuras apresentas o valor de limiar usado foi 0.1

[3.4.1] Dados Experimentais

Os dados experimentais foram extraídos do questionário MFQ30[18] e consistem de N = 14250 vetores morais sendo suas componentes relacionadas com os cinco fundamentos morais num intervalo de [0-5]. Esses dados então são normalizados para obtermos as matrizes morais experimentais. Na tabela 3.1 apresentamos algumas medidas estatísticas relativas as opiniões experimentais obtidas através do questionário.

[18] Morality Quiz/Test your Morals, Values & Ethics - Your Morals.Org

ip	n	m_e	μ_e	σ_e	*ideologia*
1	2919	0.825(5)	0.837(4)	0.084(2)	Muito Liberal
2	5604	0.877(2)	0.889(2)	0.069(2)	Liberal
3	2009	0.907(3)	0.920(3)	0.063(4)	Pouco Liberal
4	1448	0.932(3)	0.947(3)	0.056(4)	Moderado
5	879	0.964(2)	0.975(2)	0.035(3)	Pouco Conservador
6	1087	0.979(2)	0.986(1)	0.026(4)	Conservador
7	300	0.976(4)	0.987(2)	0.040(10)	Muito Conservador
6 + 7	1387	0.979(2)	0.987(1)	0.028(4)	Conservador

Tabela 3.1: retirada de [24]. O termo *ip* é o índice da filiação política dos indivíduos , m_e e μ_e são respectivamente a média e a mediana das opiniões dos respondentes, σ_e é a variância. Os erros representam 95% de confiança de reamostragens.

⌈4⌉ Resultados

> One important idea is that science is a means whereby learning is achieved, not by mere theoretical speculation on the one hand, nor by the undirected accumulation of practical facts on the other, but rather by a motivated iteration between theory and practice.
>
> — George E. P. Box, 1912-2013

A principal motivação desse trabalho é descrever, através de uma modelagem matemática de agentes interagentes, o fenômeno social do aprendizado moral e suas conexões com ideologia política, estratégia cognitiva e quantidade informação moral obtida durante a infância e adolescência. Portanto, o foco desse capítulo são os resultados computacionais que podem ser comparados diretamente com algumas evidências experimentais relativas a esses assuntos.

⌈4.1⌉ Ideologia política do agente Bayesiano

Apesar dos agentes não apresentarem ideologia política, nós podemos medir as distribuições de opiniões sobre o *Zeitgeist* $P_S(h|\rho,\beta)$ no estado estacionário da fase 2 do modelo para uma sociedade em que todos os agentes estão submetidos a um mesmo valor de β e ρ. Esta assinatura estatística pode ser comparada com a assinatura similar obtida através dos dados do questionário sobre Fundamentos Morais para cada filiação política. Isto permite encontrar a região de parâmetros do modelo que é mais similar ao perfil estatístico das opiniões de pessoas com uma determinada ideologia política.

Na figura 4.1 comparamos os histogramas de opinião, em relação ao

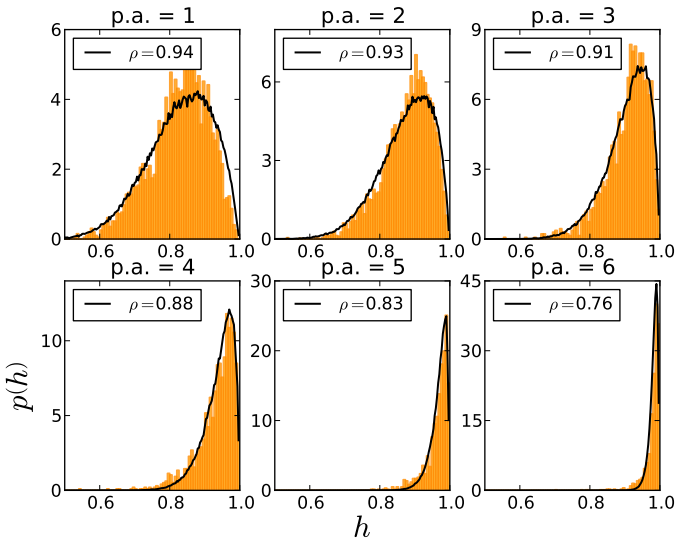


Figura 4.1: Comparação entre os histogramas de opinião, em relação ao *Zeitgeist*, obtidos através de simulação (linhas pretas) e obtidos através do questionário sobre fundamentos morais (caixas laranjas). O valor do parâmetro de pressão social usado nesses resultado foi de $\log(\beta) = 3,8$.

Zeitgeist, obtidos através de simulação (linhas pretas) e obtidos através do questionário sobre fundamentos morais (caixas laranjas). O valor do parâmetro de pressão social usado nesses resultado foi de $\log(\beta) = 3,8$. Podemos observar que existe um bom acordo entre os histogramas dos dados simulados e dos empíricos.

Já a figura 4.2 mostra como a ideologia Política (pa= 1-Muito liberal, pa=7-Muito Conservador) está correlacionada com ρ, que é uma medida da diversidade moral aprendida durante a Fase 1. Isso ocorre para uma grande variação do parâmetro β que avalia a pressão social. Percebemos claramente que quanto maior for o valor do parâmetro ρ mais os histogramas de opinião simulados se tornam similares aos histogramas experimentais de indivíduos liberais.

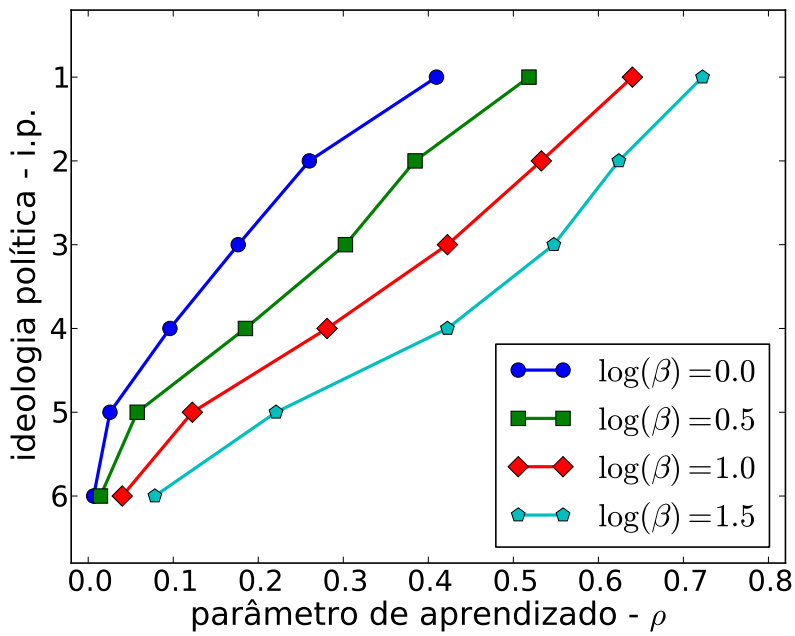


Figura 4.2: Ideologia Política (pa= 1-Muito liberal, pa=7-Muito Conservador) está correlacionada com ρ, que é uma medida da diversidade moral aprendida durante a Fase 1. Isso ocorre para uma grande variação do parâmetro β que avalia a pressão social.

[4.2] Diagrama Político

O diagrama de fase é um dos instrumentos mais úteis para a caracterização dos comportamentos em uma sociedade de agentes. Sua utilidade decorre do fato de que os contornos de fase dividem regiões[1] com comportamentos totalmente distintos.

Na figura 4.3 apresentamos o diagrama de fase nos espaço, ρ que mede a complexidade de socialização versus β que mede a pressão social. As faixas coloridas representam sociedades de agentes que podem ser estatisticamente identificadas com grupos com diferentes ideologias políticas. A linha preta indica a transição de fase de uma sociedade totalmente desorganizada (abaixo da linha) para uma organizada. A linha magenta indica os parâmetros para os quais a sociedade tem os menores tempos de correlação.

[1] No caso do diagrama de fase apresentados na figura 4.3 a linha preta é o único contorno de fase

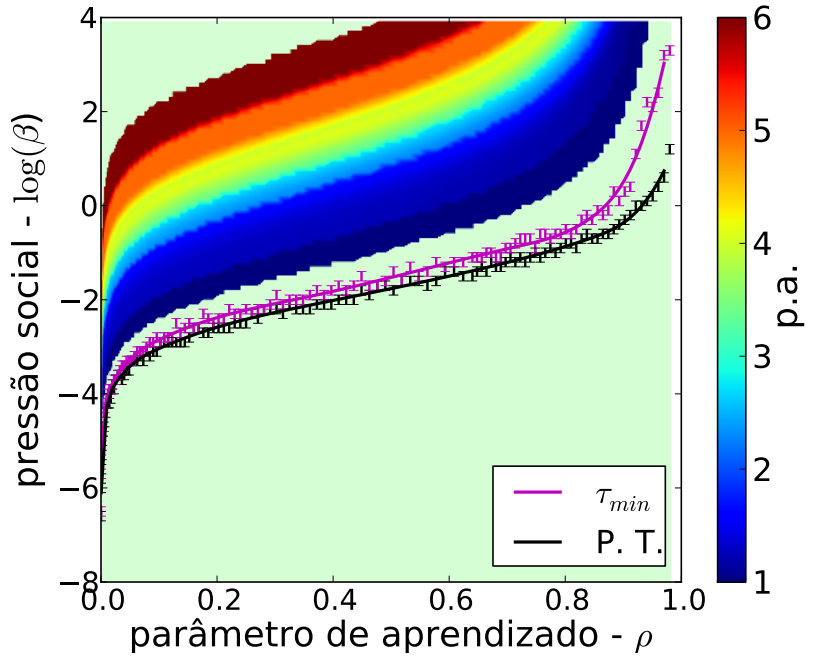


Figura 4.3: Diagrama de fase nos espaço, ρ que mede a complexidade de socialização versus β que mede a pressão social. As faixas coloridas representam sociedades de agentes que podem ser estatisticamente identificadas com grupos com diferentes ideologias políticas. A linha preta preta indica a transição de fase de uma sociedade totalmente desorganizada (abaixo da linha) para uma organizada. A linha magenta indica os parâmetros para os quais a sociedade tem os menores tempos de correlação.

[4.3] Conservadorismo e liberalismo de que?

O que o conservadorismo conserva? Se uma sociedade de agentes identificada como conservadores (pequeno ρ) se readaptassem a mudanças na sociedade mais rapidamente do que sociedades liberais, nossa teoria deveria ser sumariamente descartada. No entanto, o resultado de nossa teoria é que sociedades liberais se readaptam mais rápido que sociedades conservadoras.

Existem várias maneiras de se medir tempos de readaptação na sociedade que trazem resultados qualitativamente equivalentes para a fase ordenada do diagrama de fases.

Uma maneira de medir tempos de readaptação foi introduzida em [23], onde depois que a distribuição de probabilidade $P_S(h|\beta,\rho)$ de opiniões sobre um *Zeitgeist*, $\mathcal{Z}_{old}$ é atingida, a dinâmica é reiniciada com os agentes discutindo um novo *Zeitgeist* $\mathcal{Z}_{new}$. Com isso, um novo histograma de opinião em relação ao novo *Zeitgeist* $P_t(h)$ é medido ao

longo do tempo. A distância entre os histogramas é medida por,

$$D(t) = \sum_{h \in bins} \left(P_t(h) - P_S(h|\beta,\rho)\right)^2;$$

ou seja, somando a diferença quadrática entre as frequências de um conjunto de bins. É esperado que exista um decaimento exponencial da distância ao longo do tempo, $D(t) \propto \exp(-t/T)$, esse decaimento deve depender de um tempo de adaptação médio T que é função dos parâmetros ρ e β.

Uma segunda maneira de acessar essa informação é através do tempo de decaimento da auto-correlação da opinião dos agentes. Definimos a auto correlação como a média temporal do produto entre as opiniões médias dos agentes em dois tempos distintos subtraída do produto da média das opiniões médias nesses tempos, ou seja,

$$\begin{aligned} c(t) &= \left\langle m(t)m(t+t')\right\rangle - \left\langle m(t)\right\rangle \left\langle m(t+t')\right\rangle, \\ &= \frac{1}{t_{max}-t}\sum_{t'=0}^{t_{max}-t} m(t')m(t'+t) \\ &\quad - \frac{1}{t_{max}-t}\sum_{t'=0}^{t_{max}-t} m(t') \times \frac{1}{t_{max}-t}\sum_{t'=0}^{t_{max}-t} m(t'+t). \end{aligned} \tag{4.1}$$

Essa grandeza também apresenta um decaimento exponencial que é dado na forma

$$c(t) \propto \exp\left(-\frac{t}{\tau}\right);$$

sendo o tempo de relaxamento uma função dependente dos parâmetros de socialização e pressão social, $\tau(\beta,\rho)$.

Na figura 4.4, mostramos o resultado de medidas de tempo de correlação para diversas sociedades de agentes. Podemos notar que existe uma linha onde os tempos de auto-correlação divergem no meio de uma região com baixos tempos. Essa linha ocorre exatamente na transição de fase e é causada por um fenômeno comum em transições de fases conhecido como *relaxamento crítico*[2]. A região abaixo dessa linha onde o tempo de correlação permanece constante é a mesma onde a não existe organização social.

[2] tradução livre de: *Critical Slowing Down.*

Acima da linha de transição de fase, observa-se que para um valor

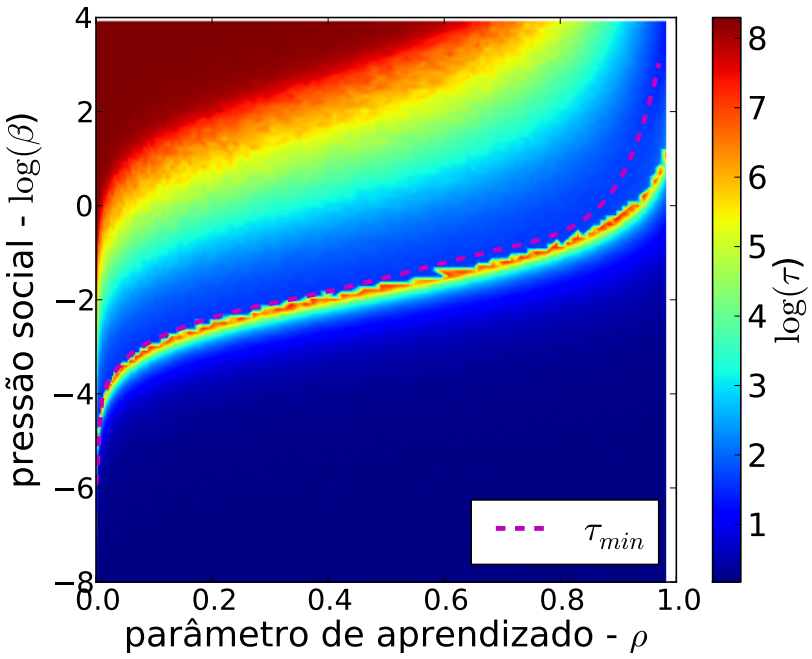


Figura 4.4: Figura com código de cores para o tempo de relaxação. Podemos observar que na linha de transição de fase o tempo de correlação diverge, esse fenômeno é conhecido como *desaceleração crítica*. Para sociedades de agentes identificadas como liberais o tempo de autocorrelação é menor do que para as sociedades de agentes identificadas como conservadores. Interessante notar que os sociedades identificadas como ultraliberais estão próximas da linha do tempo de correlação mínimo.

do parâmetro ρ fixo, à medida a pressão social aumenta o tempo de correlação τ também cresce monotonamente.

Ainda é possível verificar que os padrões de cor das região do diagrama de tempos de correlação são semelhantes às faixas de cores que identificam as sociedades de agentes com ideologias políticas através dos histogramas de opinião. Com isso, podemos concluir que a filiação política do indivíduo pode ser caracterizada pelo tempo de relaxação, e que esse cresce à medida que andamos no espectro de filiação política no sentido liberal conservador. É interessante notar que as sociedades identificadas como ultraliberais estão próximas da linha do tempo de correlação mínimo.

[4.4] Pressão Social

O parâmetro de pressão social β determina a escala de tolerância a flutuações na função de custo psicológico $\mathcal{E}$, isto é, ele determina o quão importante é para o agente estar em conformidade com seus pares. Dessa forma, o parâmetro β, no estado estacionário do modelo, informa a escala de flutuação das matrizes morais dos agentes entorno do *Zeitgeist*.

Assumindo que existe algum mecanismo, que não modelamos explicitamente, que permite ao agente fazer um julgamento do ambiente social, podemos modelar o efeito de ameaças externas ao grupo que o agente pertence considerando aumentos no parâmetro de pressão β.

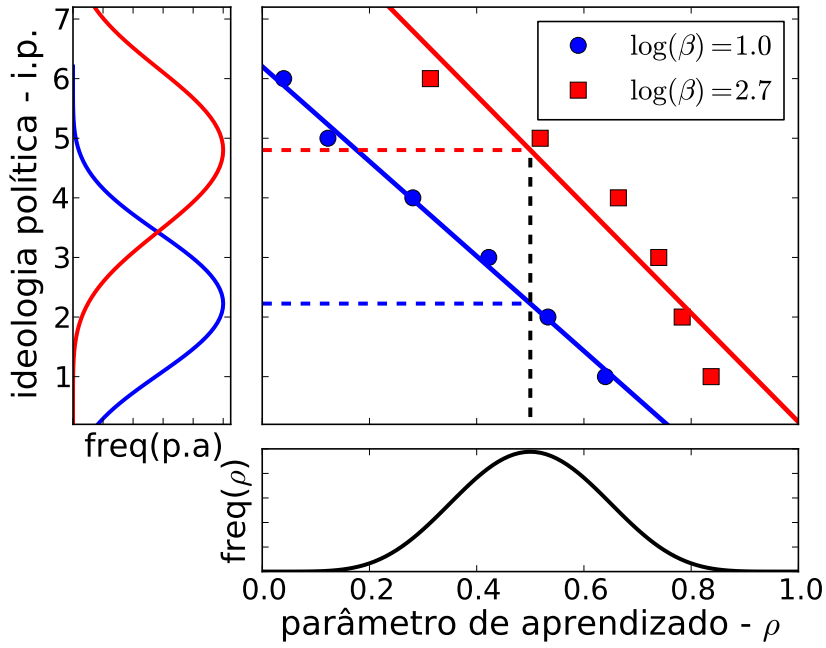


Figura 4.5: O proporção de agentes conservadores muda com a pressão social. Se a população tem uma distribuição de encontros sociais apresentada na parte de baixo da figura a distribuição da filiação política resultante varia com a pressão social, como é mostrado na figura a direita.

Supondo uma distribuição de probabilidade para o número de interações sociais entre os agentes na fase 1, ou seja, uma distribuição de probabilidade fixa para o parâmetro ρ (figura 4.5, gráfico em baixo), a distribuição de ideologia política na população irá variar de acordo com o parâmetro β (gráfico a esquerda). Como podemos ver pela fi-

gura 4.5, para uma sociedade com uma alta pressão social os agentes se distribuirão com mais concentração na parte conservadora do espectro político, e com menos pressão social se concentra em torno da parte mais liberal do espectro.

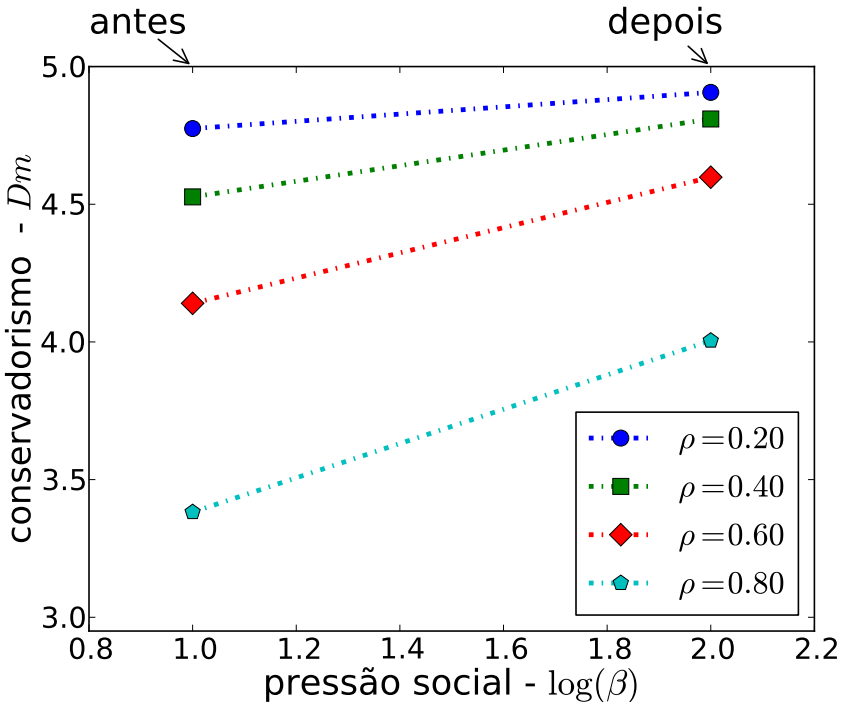


Figura 4.6: O número efetivo de dimensões morais para dois valores de pressão social, antes e depois de uma ameaça. Se ameaças acarretam em aumento da pressão social, a assinatura estatística de liberais se torna mais parecida com a de conservadores.

Outra maneiras de percebermos esse mesmo efeito é olhando para o índice de conservadorismo, definido pelo número efetivo de dimensões morais usadas na sociedade, que nada mais é que número de dimensões morais multiplicado pela opinião média dos agentes em relação ao *Zeitgeist*, $Dm = 5m$. Na figura 4.6 observamos o efeito de crescimento da dimensão moral efetiva da sociedade com diferentes valores de ρ no caso de um aumento na pressão social devido a uma ameaça.

[4.5] Parceiros sociais

Em nosso modelo, além dos parâmetros ρ de aprendizagem do agente e β da pressão entre pares na sociedade, outro fator importante que

deve ser estudado é a dependência do comportamento com o número médio de parceiros sociais, k, dos agentes[3].

Na figura 4.7 é apresentado a dependência das opiniões médias dos agentes, m, e do tempo de adaptação da sociedade, τ, em função do número médio de vizinhos dos agentes, para sociedades com diferentes pressões sociais ($\log(\beta) = -1, 0, 1$) mas com agentes com a mesma valor para o parâmetro da encontros sociais da fase 1 ($\rho = 0.9$).

[3] Os resultados a seguir foram obtidos com os agentes dispostos sobre redes de Barabasi-Albert [5].

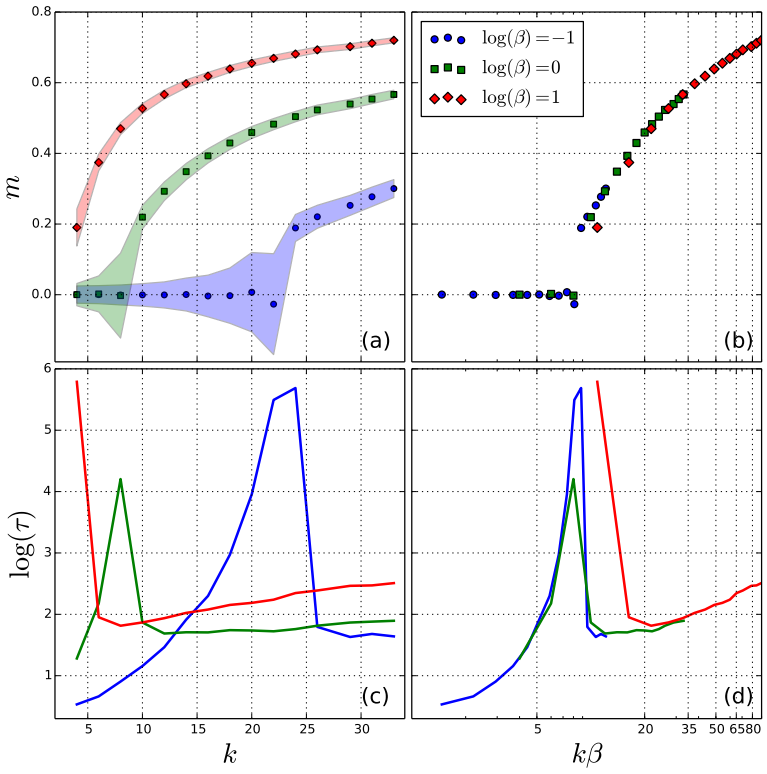


Figura 4.7: Dependência com o número de parceiros sociais e fator de escala para pressão social. À esquerda são apresentados, para diferentes valores de pressão social, β, a opinião média dos agentes (em **a**) e o logaritmo do tempo de adaptação (em **c**) em função do número de médio de vizinhos de um agente na sociedade. Observa-se a existência de uma transição de fase entre os estados de sociedade desordenados e ordenados à medida que o número médio de vizinhos cresce. As regiões de contorno da figura **a** representam o erro da média das opiniões dos agentes. Nas figuras à direita, em escala logarítmica, é apresentado novamente m e $\log(\tau)$ só que em função do número de vizinhos médios vezes a pressão social, β. Percebemos claramente que devido ao fator de escala, as curvas das opiniões médias e tempos de adaptação para diferentes valores de pressão social se colapsam em uma única curva. As simulações foram feitas com 400 agentes de tendência corroborativa $\rho = 0.9$ em uma rede Barabasi-Albert.

Podemos observar, a partir da figura 4.7(a), a existência de uma transição de fase entre uma sociedade desordenada ($m = 0$) para uma sociedade ordenada ($m > 0$) à medida que o número médio de vizinhos, k, cresce. Além disso, percebemos que quanto maior é o valor da pressão social β menor é a quantidade de parceiros sociais necessários para a sociedade entrar em conformidade. Nessa mesma figura,

as regiões de contorno representam o erro da média das opiniões dos agentes, que, como o esperado, apresentam um comportamento divergente na transição de fase. A transição de fase entre os estados desordenados e ordenados também pode ser confirmado pelo comportamento divergente do tempo de adaptação da sociedade como é mostrado na figura 4.7(b) onde é apresentado o logaritmo do tempo de adaptação τ em função do número de médio de vizinhos de um agente na sociedade.

Uma interpretação mais interessante para a dependência do modelo com o numero de parceiros sociais pode ser feita ao observamos os gráficos (b) e (d) da coluna a direita da figura 4.7, é apresentado novamente os valores da opinião média da sociedade m e seu tempo de adaptação $\log(\tau)$ em função do número médio de parceiros sociais vezes a pressão social, $k\beta$. Percebemos claramente, que, devido a essa mudança de escala, as curvas das opiniões médias e tempos de adaptação para diferentes valores de pressão social se colapsam em uma única curva. Isso implica que, em nosso modelo, a pressão social β e o número efetivo de parceiros sociais k, apesar de natureza distintas, influenciam o comportamento da sociedade de agentes de maneira equivalente.

⌈4.6⌉ Detalhes computacionais

Todos os resultados de simulação descritos nesse capítulo, exceto a figura 4.7, assim como os gráficos 3.2 e 3.3 do capítulo anterior, foram obtidos através da simulação de sociedade com 400 agentes dispostos sobre uma rede do tipo Barabasi-Albert com um número médio de 20 vizinhos. Escolhemos esse número de vizinho pois evidências experimentais [4] sugerem que o número de parceiros sociais relevantes de uma pessoa está entre 18 e 22. A dependência com a estrutura da rede foi estudada no contexto do modelo proposto por Caticha e Vicente[24] no apêndice C.2 e na dissertação de mestrado de Jericó [5]. Para um revisão sobre grafos complexos e suas aplicações nos mais diversos contextos recomendamos ao leitor a referência [5].

[4] M. Trusov, A. Bodapati, and R. Bucklin. Determining influential users in internet social networks. *Journal of marketing research*, XLVII(August):643–658, 2010

[5] J. P. Jericó. *Aplicações de Mecânica Estatística a Sistemas Sociais : Interação e Evolução Cultural*. Mestrado, Universidade de São Paulo, 2012

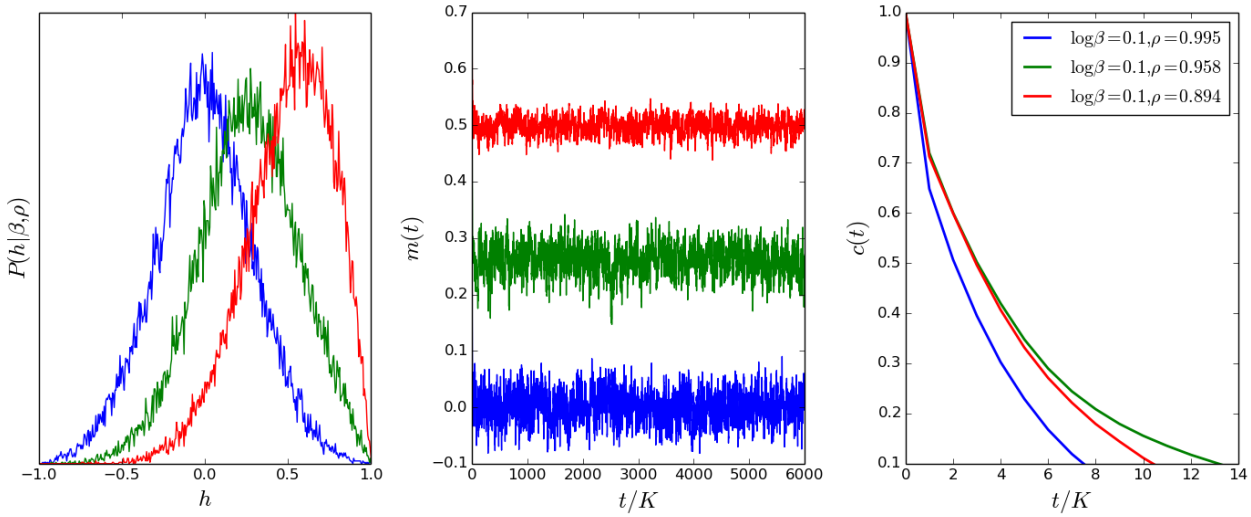


Figura 4.8: Figura com resultados de situação de 3 experimentos de Monte Carlo de uma sociedade de 400 agentes ($K = 400$) em uma rede Barabasi-Albert com 20 vizinho ($n = 20$) para diferentes valores de pressão social e parâmetro de socialização. À esquerda é apresentado o histograma $P(h|\beta,\rho)$, ao centro opinião média dos agentes $m(t)$ em função do tempo, à direita a auto-correlação temporal das opiniões médias $c(t)$.

Cada ponto dos gráficos apresentados foram resultados de médias de pelo menos 800 configurações de Monte Carlo independentes, dependendo do tempo de correlação da simulação. As simulações foram feitas em 33 computadores de quatro núcleos e velocidade de processamento entre 2.4 e 2.8 Ghz. Foram gastos aproximadamente 8 dias de simulação contínua em todos os núcleos.

Na figura 4.8 é apresentado o histograma de opiniões $P(h|\beta,\rho)$, opinião média dos agentes $m(t)$ em função do tempo, a auto-correlação temporal das opiniões médias dos agentes $c(t)$, de uma rodada de Monte Carlo para diferentes valores de pressão social e parâmetro de socialização.

[5] Análise multicultural

> For if society lacks the unity that derives from the fact that the relationships between its parts are exactly regulated, that unity resulting from the harmonious articulation of its various functions assured by effective discipline and if, in addition, society lacks the unity based upon the commitment of men's wills to a common objective, then it is no more than a pile of sand that the least jolt or the slightest puff will suffice to scatter.
>
> — David Émile Durkheim, 1858-1917

Uma das perguntas mais importantes que podem ser feitas é o quão geral é a teoria de aprendizado moral que desenvolvemos durante a tese. Se ela só é valida para os Estados Unidos da América, ou se é geral o suficiente para descrever o aprendizado moral de grande parte das sociedades modernas. Esse tema começou a ser tratado após a conclusão da tese, quando obtivemos o conjunto de dados completo do questionário da Teoria dos Fundamento Morais[2].

Para abordarmos essa questão, iremos adotar o modelo proposto por Caticha e Vicente[1], por ter a vantagem de ser um sistema Hamiltoniano com um potencial de interação mais simples que a do modelo de agentes Bayesianos. Isso permite, como demonstrado em [112], que seja feita uma análise de campo médio, que foi melhor desenvolvida em [119]. Usando os resultados analíticos fornecidos pelo campo médio, conseguimos analisar dados experimentais sem que seja necessário a comparação entre os histogramas experimentais e os produzidos a partir de simulação.

[1] N. Caticha and R. Vicente. Agent-Based Social Psychology: From Neurocognitive Processes To Social Data. *Advances in Complex Systems*, 14(05):711, 2011

A pesquisa social multicultural estuda a variação de comportamento

humano, e como ele está relacionado ao seu contexto cultural. De acordo com Berry *et al.*[2], os estudos multi culturais podem ser divididos em três categorias ou orientações: **relativista**, **absolutista** e **universalista**[3].

A orientação relativista busca evitar qualquer traço de etnocentrismo tentando entender as pessoas em "seus próprios termos", ou seja, ela tenta entender as pessoas usando as "categorias e valores" da cultura do indivíduo. De maneira geral, a pesquisa de orientação relativista supõe que variações de comportamentos entre indivíduos se devem quase que exclusivamente a fatores culturais. Além disso, pesquisadores relativistas não mostram muito interesse em similaridades entre culturas, sendo que as diferenças culturais são tipicamente interpretadas de forma qualitativa. Em contraste, pesquisadores absolutistas apresentam pouca preocupação com problemas de etnocentrismo assumindo que as pessoas são basicamente iguais em todos os lugares, e que quando diferenças existem elas podem ser medidas quantitativamente. A posição universalista supõe que existem processos psicológicos básicos que são comuns a todos os seres humanos, no entanto, suas manifestações são provavelmente influenciadas pela cultura. Assim sendo, o foco da pesquisa universalista está em identificar quais fatores culturais e biológicos influenciam o comportamento da sociedade e de seus indivíduos.

Apesar de abordagens mais matemáticas para a descrição dos problemas sociais darem a impressão de que é adotada uma perspectiva absolutista, acreditamos que nosso modelo é melhor entendido através da perspectiva universalista, já que estamos tentando, de forma esquemática, descrever o resultado da interação entre componentes biológicos e culturais através da modelagem de agentes, incorporando uma série de características observadas nos mais diversos ramos da pesquisa social (como discutido ao longo do primeiro e segundo capítulo do texto).

A discussão sobre essa análise será apresentada na seguinte ordem, primeiramente, faremos uma breve introdução ao modelo de agentes proposto por Caticha e Vicente[24][4]. Em seguida, descreveremos a

[2] J. W. Berry, Y. H. Poortinga, M. H. Segal, and P. R. Dasen. *Cross-cultural psychology: Research and applications*. Cambridge University Press, Cambridge, second edition, 2002

[3] tradução livre de: *relativism*, *absolutism* e *universalism*.

[4] Uma descrição mais detalhada do modelo juntamente como uma série de resultados computacionais é feita no apêndice C.

aproximação de campo médio para esse modelo e introduziremos uma medida de máxima verossimilhança que permite comparar os parâmetros do modelo com os dados experimentais. Por fim, usando essa teoria, analisaremos os dados das matrizes morais de indivíduos de 23 países e compararemos o resultado com o obtido por outro estudo multicultural [44] que visa estabelecer (através de uma escala obtida por um questionário) o quão *rígida* ou *flexivel* é a cultura de um país.

⌈5.1⌉ Agentes com tendência corroborativa

A dinâmica de aprendizado moral para o modelo de agentes proposto em [24] é equivalente e dinâmica da segunda fase de aprendizado do modelo de agentes Bayesianos. No modelo de Caticha e Vicente [24], não existe uma primeira fase de aprendizado moral onde os agentes irão aprender a lidar com novas informações em uma fase adulta. Um dos principais resultados obtido em [24] foi a conexão entre a estratégia cognitiva dos agentes e o padrão estatístico dos histogramas de opinião moral de conservadores e liberais.

Nesse modelo, assim como na segunda fase do modelo de agentes Bayesianos, um agente i muda sua matriz moral ao discutir um assunto na direção do *Zeitgeist* se isso diminuir sua energia, ou custo psicológico, de interação com parceiros sociais. Caso o custo cresça, o agente ainda tem uma probabilidade de mudar de opinião que diminui exponencialmente com a variação do custo psicológico vezes um fator de escala α. O fator de escala[5], o qual denominamos de *pressão social*, avalia quais as amplitudes de mudanças de opiniões são aceitáveis na sociedade, ou seja,

[5] Nesse capítulo, o termo de pressão social do modelo é chamado de α para diferenciá-lo da pressão do modelo Bayesiano, β; apesar de ambos os termos terem a mesma interpretação.

$$\boldsymbol{\omega}_i \to \boldsymbol{\omega}_i' \text{com prob.} p = \min\{1, \exp(-\alpha \Delta \mathcal{E}_\delta)\},$$

sendo,

$$\Delta \mathcal{E}_\delta = \sum_{k \in \mathcal{V}_i} \left[\mathcal{E}_\delta(h_i', h_k) - \mathcal{E}_\delta(h_i, h_k) \right],$$

onde $\mathcal{V}_i$ são os parceiros sociais do agente i e $h_i = \mathcal{Z} \cdot \omega_i$. A energia de interação entre dois parceiros sociais introduzida por Caticha e Vicente

[24] é definida como,

$$\mathcal{E}_\delta\left(h_i, h_j\right) = -\frac{1+\delta}{2} h_i h_j + \frac{1-\delta}{2}|h_i||h_j|. \tag{5.1}$$

Onde $\delta \in [0,1]$ é o parâmetro que define a estratégia cognitiva do agente. O custo psicológico da interação entre agentes, assim como o parâmetro δ, pode ser mais facilmente interpretada olhando para a função de modulação do aprendizado do agente, que é gerada a partir do gradiente da energia[6]. Assim sendo, a função de modulação definida a partir da energia de interação 5.1 é dada por,

$$F_\delta\left(h_i, \sigma_j\right) = \begin{cases} 1 & \text{caso } h_i h_j < 0, \\ \delta & \text{caso } h_i h_j > 0. \end{cases} \tag{5.2}$$

Através da equação da função de modulação 5.2, e também pela figura 5.1, vemos que o parâmetro δ define a estratégia cognitiva do agente, pois ele caracteriza a importância dada por um agente i a uma opinião concordante de um outro agente j em uma interação social ($h_i h_j > 0$).

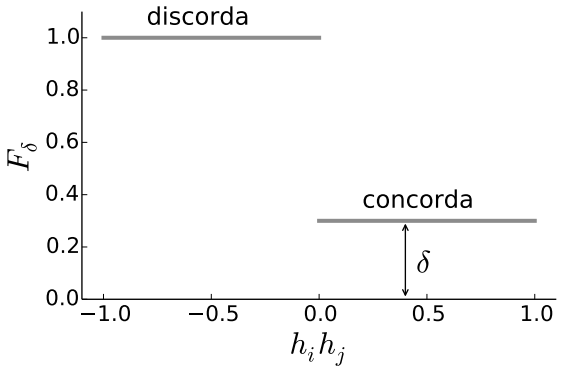


Figura 5.1: Função de Modulação de [24]. O parâmetro δ representa a tendência que o agente tem de entrar em conformidade, ou seja, quanto maior for o valor de δ mais o agente irá aprender com um assunto que ele é concor-

Sendo assim, como δ representa a importância das opiniões concordantes essa grandeza é denominada de *tendência corroborativa* do agente. Por outro lado, o grau de importância à opiniões discordantes ($h_i h_j < 0$) relativamente às concordantes é dado por $1 - \delta$. Dessa maneira, podemos interpretar essa grandeza como um quantificador do aprendizado relacionado a informações desconhecidas ou conflitantes.

Como sugerido por Caticha e Vicente[24] o parâmetro δ que de-

[6] Ou seja,

$$F_\delta(h_i, h_j) = \frac{\partial}{\partial h_i}\mathcal{E}_\delta(h_i, h_j),$$

de maneira equivalente,

$$\mathcal{E}_\delta(h_i, h_j) = \int_0^{h_i} \mathrm{d}h F_\delta(h, h_j).$$

No contexto pressão social infinita a dinâmica de Monte Carlo equivale a uma dinâmica de descida de gradiente da energia de interação.

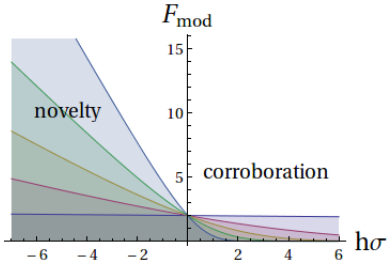


Figura 5.2: Dependência da função de modulação com a quantidade de troca de informação medido através de ρ. Quanto mais informação é processada maior é a amplitude da função de modulação para informações.

fine a estratégia cognitiva do agente pode ser usado para mimetizar as diferenças entre as estratégias cognitivas de pessoas liberais e conservadores como foi evidenciado no experimento de [7], e discutido com mais profundidade na seção 2.3, onde pessoas liberais apresentam uma maior ativação cerebral para informações conflitantes relativa à informação esperadas do que pessoas conservadoras.

Como discutido no capítulo 3, a estratégia cognitiva do agente Bayesiano muda à medida que ele ganha experiencia na fase 1. Com isso, à medida que o parâmetro ρ, que mede a quantidade de informação adquirida na fase 1, cresce, a função de modulação do agente Bayesiano passa a dar mais importância à assuntos para os quais existem discordância (ver figura 5.2). Além disso, como demonstrado no apêndice C.6, no limite termodinâmico, onde o número de assuntos discutidos entre parceiros sociais tende ao infinito podemos interpretar que a tendência corroborativa efetiva do agente Bayesiano é dada por $\tilde{\delta} \approx 1 - \rho$.

⌈5.2⌉ Campo Médio

A aproximação de campo médio é feita ao maximizarmos a entropia relativa entre uma distribuição de probabilidade do tipo $P_{cm} = \prod_i Q_i(\boldsymbol{\omega}_i)$ em relação a distribuição de Gibbs[7], $P_G \propto \exp -\alpha\mathcal{H}$, onde a entropia relativa entre as duas distribuições é dada por,

[7] M. Opper and D. Saad. *Advanced mean field methods: Theory and practice*. The MIT Press, London, England, 2001

$$\begin{aligned} S\left[P_{cm}||P_G\right] &= -\int \prod_i d\mu\left(\boldsymbol{\omega}_i\right)\left\{P_{cm}\log\frac{P_{cm}}{P_G} - \lambda\left(P_{cm}-1\right)\right\} \\ &= -\int d\mu\left(\boldsymbol{\omega}_i\right) Q_i \log Q_i + \lambda Q_i, \\ &\quad - \alpha\sum_{(i,j)}\int d\mu\left(\boldsymbol{\omega}_i\right) d\mu\left(\boldsymbol{\omega}_j\right) Q_i Q_j V_{ij} + cte, \end{aligned} \tag{5.3}$$

onde $d\mu(\boldsymbol{\omega})$ é o elemento de área de uma hiperesfera. A maximização da entropia acontece quando $\frac{\delta S[Q||P_G]}{\delta Q_i} = 0$, de onde segue[8],

[8] No caso do modelo de Ising essa procedimento resulta na aproximação de campo médio de Curie-Weiss[89]

$$Q_i \propto \exp\left(-\alpha\sum_{j\in viz(i)}\int d\mu\left(\boldsymbol{\omega}_j\right) Q_j V_{ij}\right).$$

Devido ao formato do Hamiltoniano 5.1, podemos definir os seguintes parâmetros de ordem,

$$m = \int d\mu\left(\boldsymbol{\omega}_j\right) Q_j \boldsymbol{\omega}_j \cdot \mathcal{Z},$$
$$r = \int d\mu\left(\boldsymbol{\omega}_j\right) Q_j |\boldsymbol{\omega}_j \cdot \mathcal{Z}|,$$

de forma que a Hamiltoniana de campo médio será dada por,

$$\int d\mu\left(\boldsymbol{\omega}_j\right) Q_j V_{ij} = -\frac{1+\delta}{2} h_i m + \frac{1-\delta}{2}|h_i| r.$$

Com isso, a distribuição de probabilidade pela aproximação de campo médio será dada por,

$$\begin{aligned} P_{cm}\left(\boldsymbol{\omega}|\alpha,\delta,\{m,r\}\right) &= \prod_i \frac{1}{Z} \exp\alpha\left(-\frac{1+\delta}{2} h_i m + \frac{1-\delta}{2}|h_i| r\right), \\ &= \prod_i \frac{1}{Z} B(h_i|\delta, k\alpha, m, r), \end{aligned} \tag{5.4}$$

sendo obtida a partir da solução das seguintes equações auto consistentes,

$$m = \frac{1}{Z}\int d\mu\left(\boldsymbol{\omega}\right) B(h|\delta, k\alpha, m, r) h, \tag{5.5}$$
$$r = \frac{1}{Z}\int d\mu\left(\boldsymbol{\omega}\right) B(h|\delta, k\alpha, m, r)|h|, \tag{5.6}$$
$$Z = \int d\mu\left(\boldsymbol{\omega}\right) B(h|\delta, k\alpha, m, r). \tag{5.7}$$

Considerando a direção do *Zeitgeist* como o eixo de simetria segue que,

$$P\left(h|\alpha,\delta,\{m,r\}\right) = \int d\mu(\boldsymbol{\omega})\delta(\mathcal{Z}\cdot\boldsymbol{\omega} - h) Q\left(\boldsymbol{\omega}|\alpha,\delta,\{m,r\}\right). \tag{5.8}$$

Experimentalmente, estamos interessados na região de parâmetros onde a população se organiza com grande probabilidade na direção do *Zeitgeist*, ou seja, quando $h_i > 0$, para quase todo agente i. Nessa situação $m \approx r$, e assim a distribuição de campo médio será dada por,

$$P_{cm}(h|\delta, k\alpha, \{m\}) \approx \frac{\gamma^2}{2}\left(1-h^2\right) e^{-\gamma(1-h)}, \tag{5.9}$$

onde $\gamma = m\delta k\alpha$.

[5.2.1] Estimação de máxima verossimilhança

Considerando um conjunto de dados de opinião, $D = \{h_1, \ldots, h_n\}$, independentes e identicamente distribuídos a partir da probabilidade de campo médio (5.9), teremos que o logaritmo da verossimilhança para esse conjunto de dados será dada por,

$$L(\gamma) = \sum_{i=1}^{n} \log P_{cm}(h_i|\delta, k\alpha, \{m\}) = \sum_{i=1}^{n} \log P_{cm}(h_i|\gamma). \tag{5.10}$$

Com isso, podemos estimar o valor do parâmetro γ que maximiza a probabilidade da amostra ter sido sorteada,

$$\frac{dL(\gamma)}{d\gamma} = \sum_i (-1 + \frac{2}{\gamma} + h_i) = 0. \tag{5.11}$$

Logo, o valor do parâmetro γ que maximiza a verossimilhança é dado por,

$$\begin{aligned} \frac{2}{\gamma} &= 1 - \frac{1}{n}\sum_i h_i, \\ &= 1 - \overline{m}. \end{aligned} \tag{5.12}$$

Assumindo que o número de amostras é muito grande ($n \to \infty$), pela lei dos grandes números[9] a média amostral de uma variável aleatória tende à esperança dessa variável e com isso temos $\overline{m} \to m$. Lembrando que $\gamma = m\delta k\alpha$, obtemos a partir do estimador de máxima verosimilhança em (5.12) que

[9] M. H. DeGroot. *Probability and statistics*. Addison-Wesley, 1989; and L. Wassermann. *All of statistics: A Concise Course in Statistical Inference*. Springer-Verlag, 2003

$$m(1 - m) = \frac{2}{k\alpha\delta}. \tag{5.13}$$

Como assumimos m estritamente positivo, temos que a opinião média dos agentes em função do número médio de vizinhos, pressão social e tendência corroborativa na aproximação de campo médio quando $m \approx r$ é

$$m(k\alpha, \delta) = \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{2}{k\alpha\delta}}. \tag{5.14}$$

De fato, a qualidade da aproximação de campo médio para a regres-

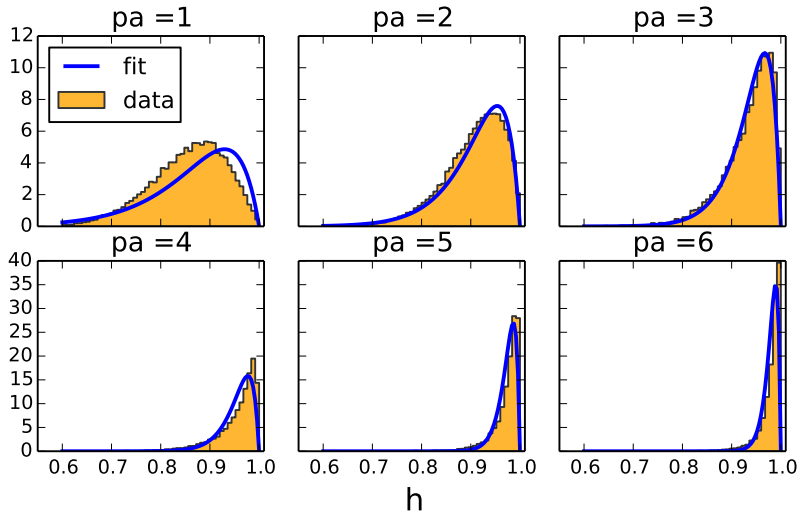


Figura 5.3: Em amarelo são apresentados os histogramas das opiniões morais para grupos de pessoas com diferentes ideologias políticas de 113 mil respondentes dos Estados Unidos da América. As linhas azuis são as distribuição de probabilidade de campo médio estimada a partir da maximização da verossimilhança.

são dos histogramas de opinião experimentais pode ser verificada pela figura 5.3, onde, em amarelo, são apresentados os histogramas das opiniões morais de 113 mil respondentes dos Estados Unidos da América com diferentes ideologias políticas, e as linhas azuis as distribuições de campo médio estimada a partir da maximização da verossimilhança.

Além disso, através da figura 5.4, onde são apresentados os histogramas de opiniões de respondentes do Brasil e do Japão, para diferentes ideologias políticas, podemos notar que a distribuição de campo médio também é uma boa aproximação para a distribuição de opiniões morais de indivíduos de outras culturas.

É importante notar, a partir da equação 5.14, que a metodologia da maximização da verossimilhança, usando a expressão da distribuição de campo médio 5.9, só permite estimar o parâmetro $\gamma = k\alpha\delta$, não estabelecendo assim uma dependência direta com cada um dos parâmetros k, α e δ separadamente. Sendo assim, na próxima seção, iremos introduzir uma metodologia para calcularmos esses parâmetros usando um conjunto de dados com a resposta do questionário dos fundamentos morais para respondentes de vários países.

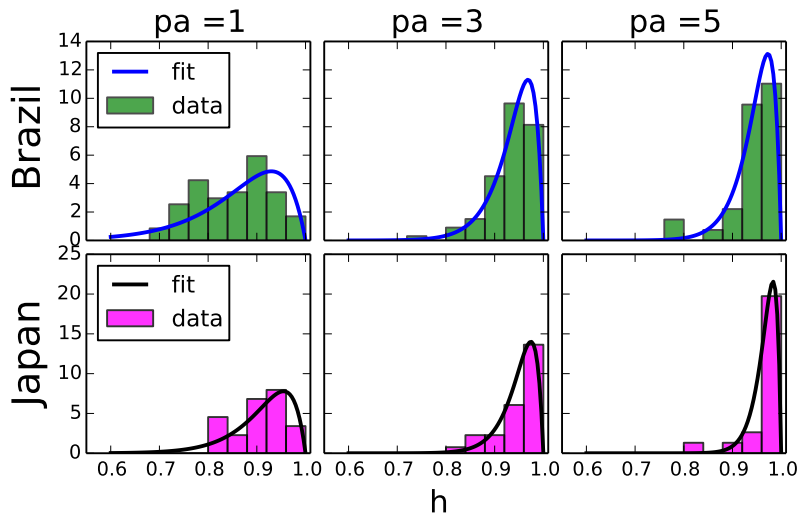


Figura 5.4: Histogramas de opiniões de respondentes do Brasil (figuras de cima) e do Japão (figuras inferiores), para diferentes ideologias políticas. Percebesse que existe uma maior dispersão nas opiniões dos respondentes Brasileiros em relação aos Japoneses com a mesma ideologia política.

[5.3] Comparação com índice de *rigidez/flexibilidade*

Um dos papéis da pesquisa de sociedades é entender a similaridade e diferenças entre diferentes culturas. Nesse contexto o trabalho publicado recentemente pela revista *Science*[44] coletou dados de 33 nações que ilustram a diferença entre culturas *rígidas*[10] e culturas *flexíveis*[11]. De acordo com os autores, culturas rígidas têm varias normas sociais fortes e pouca tolerância para comportamentos desviantes, enquanto ao contrário, culturas flexíveis apresentam normas sociais fracas e grande tolerância para comportamentos desviantes.

Através de um questionário com dezenas de perguntas, que foram respondidas por aproximadamente 200 habitantes de cada país, o autores tentaram quantificar o quanto um país é *rígido* ou *flexível*. A pergunta típica é: De 0 a 6 (0 totalmente discorda, 6 totalmente concorda) quanto você concorda com as asserções a seguir? A asserções são, "Existem muitas normas sociais que as pessoas tem de conformar-se nesse país", "Nesse pais, se alguém procede de maneira inapropriada as pessoas iram desaprovar fortemente" e "Pessoas nesse país quase sempre obedecem as normas sociais"[12]. Além disso, os participantes deveriam julgar o quão apropriado é um comportamento em alguma situação cotidiana, onde exemplos de comportamentos são: comer, rir,

[10] tradução livre de:*tight cultures*

[11] tradução livre de:*loose cutures*

[12] Tradução livre de: *"There are many social norms that people are supposed to abide by in this country,""In this country, if someone acts in an inappropriate way, others will strongly disapprove", and "People in this country almost always comply with social norms."*

chorar, cantar, etc e exemplos de situação são: na biblioteca, no banco, elevador, funeral, etc. Assim, quanto maior o valor do índice calculado pelo questionário maior é a *rigidez* do país.

Uma das conclusões mais importantes desse trabalho é que o quão *rígido/flexivel* é o país está relacionado com um sistema complexo e integrado de ameaças à sociedade de origem ecológica e humana. Por exemplo, países mais rígidos apresentam, em média, maior densidade populacional e maiores projeções de crescimento populacional. Estes apresentam maior escassez de recursos naturais incluindo menor produção de alimentos e acesso a água potável. Além disso, países mais rígidos enfrentam mais desastres naturais como enchentes, ciclones tropicais, estiagens. Quando comparado com países mais *flexíveis*, países *rígidos* apresentam uma maior incidência de conflitos seus vizinhos e também maior incidência de pestilências.

Em nosso modelo, o parâmetro de pressão social α determina o quão importante é para o agente estar em conformidade com seus pares, informando assim a escala de flutuação das matrizes morais dos agentes em torno do *Zeitgeist*. Portanto, esperamos que em países onde a sociedade tem menor tolerância para comportamentos desviantes os histogramas de matrizes morais mais serão coesos ou menos desviantes em relação ao *Zeitgeist*.

Usando os dados completos do questionário dos fundamentos morais[13], que contem as respostas de mais de 120 países e aproximadamente 300 mil respondentes, podemos, para cada país calcular o *Zeitgeist*, matrizes morais e opiniões morais usando o mesmo procedimento descrito na secção 3.4. Dentre os 120 países para os quais obtivemos dados, selecionamos 23 que tinham pelo menos 200 respondentes e também faziam parte da lista de países analisados no artigo[44]. Estes países, assim como os seus índices de *rigidez/flexibilidade* estão listados na tabela 5.1.

[13] Morality Quiz/Test your Morals, Values & Ethics - Your Morals.Org

Na figura 5.5 são apresentadas as opiniões médias dos respondentes m por ideologia política de cada país em função do índice *rigidez/flexibilidade*. Os números ao lado da ideologia política na legenda são as covariâncias entre a opinião média e índice de *rigidez/flexibilidade*.

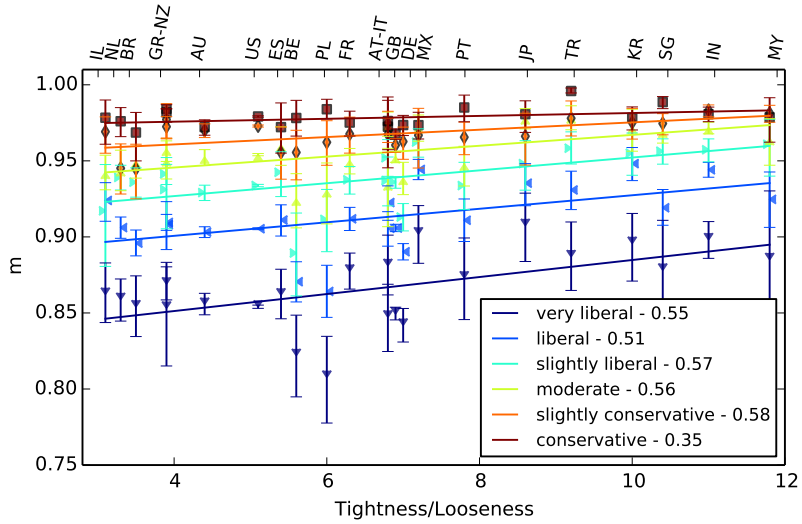


Figura 5.5: Opinião média dos respondentes m por ideologia política de cada país em relação ao índice *rigidez/flexibilidade*. Os números ao lado da ideologia política na legenda são as covariâncias entre a opinião média e índice *rigidez/flexibilidade*. Lembrando que quanto maior o índice *Tightness/Looseness* mais rígida é a cultura.

É marcante notar que de fato, a coesão da matriz moral de indivíduos de uma mesma ideologia política em média cresce com o índice de *rigidez/flexibilidade* do país. Esse efeito também pode ser visto, na figura 5.4, pela maior da dispersão das opiniões dos histogramas do Brasil (*rigidez/flexibilidade* = *3.5*) quando comparado com a dispersão dos histogramas de opinião dos respondentes de mesma ideologia política do Japão (*rigidez/flexibilidade* = *8.6*).

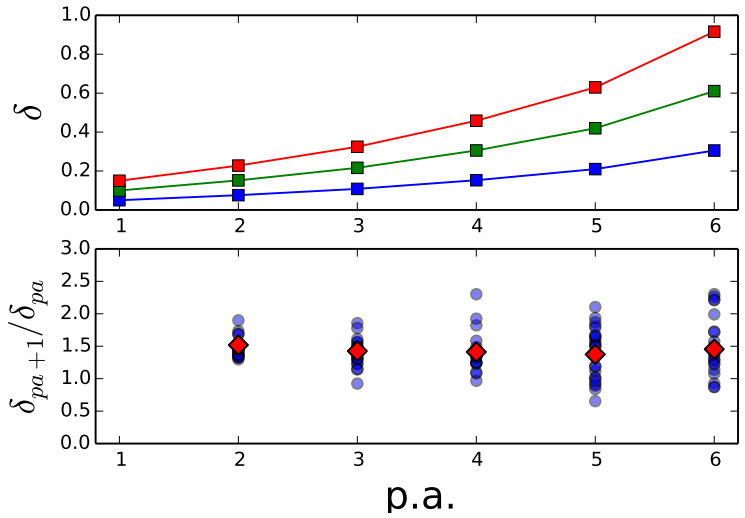


Figura 5.6: Na figura superior é apresentada a tendência corroborativa em função da filiação política. Na figura de baixo os pontos azuis representam proporção das tendências corroborativas entre pessoas de ideologias politicas consecutivas, na escala liberal conservador, para cada país. Os pontos vermelhos são as medianas das proporções das tendências corroborativas.

Assumindo que a pressão social (α_c) e o número médio de parceiros sociais (k_c) é constante dentro do país e que a tendência corroborativa dos indivíduos (δ_{pa}) muda com a ideologia política mas é independente do seu país de origem , teremos que a equação 5.13 pode ser rescrita como,

$$k_c \alpha_c \delta_{pa} = \frac{2}{m_{c,pa}\left(1 - m_{c,pa}\right)} \tag{5.15}$$

onde c é o índice do país e pa é o índice da ideologia política. Portanto, a razão entre as tendências corroborativas de grupos com índices de filiação políticas consecutivas pode ser estimado através da expressão,

$$\frac{\delta_{pa+1}}{\delta_{pa}} = \frac{1}{C} \sum_c \frac{\overline{m}_{c,pa}\left(1 - \overline{m}_{c,pa}\right)}{\overline{m}_{c,pa+1}\left(1 - \overline{m}_{c,pa+1}\right)}, \tag{5.16}$$

onde C é o número de países, $\overline{m}_{c,pa}$ indica o valor estimado das opiniões de pessoas com a mesma ideologia política em um país.

No gráfico inferior da figura 5.6 nos círculos azuis é apresentada a proporção das tendências corroborativas entre grupos de ideologias politicas consecutivas para cada país. Com isso, a proporção entre as tendências corroborativas de grupos de filiação política são indicadas no gráfico pelos losangos vermelhos[14]. É marcante observar a partir dessa figura, que $\delta_{pa+1}/\delta_{pa} \approx 1.5$ independentemente das filiações políticas. Já o gráfico superior da figura 5.6 apresenta curvas para a tendência corroborativa em função da filiação política tomando três valores distintos para δ da filiação política mais liberal ($pa = 1$).

[14] As proporções entre as tendências corroborativas, δ, foram calculadas usando medianas já que para distribuições de probabilidade simétricas elas fornecem uma medida mais robusta para a estimativa da média.

Portanto, ao definirmos os valores das tendências colaborativas médias de indivíduos com a mesma filiação política, podemos estimar o valor da pressão social vezes o numero de parceiros sociais de cada país através da expressão,

$$k_c \alpha_c = \frac{1}{6} \sum_{pa=1}^{6} \left[\delta_{pa} \overline{m}_{c,pa}\left(1 - \overline{m}_{c,pa}\right)\right]^{-1} \tag{5.17}$$

Na figura 5.7 é apresentada a relação entre a pressão social do modelo de agentes e o *tightness score*[15]. A linha vermelha representa o melhor ajuste linear ponderado pelo erros dos pontos, onde a região em azul representa o intervalo de 68% de confiança para essa regres-

[15] As barras de erros foram calculadas a partir de 95% de confiança do *bootstrap*

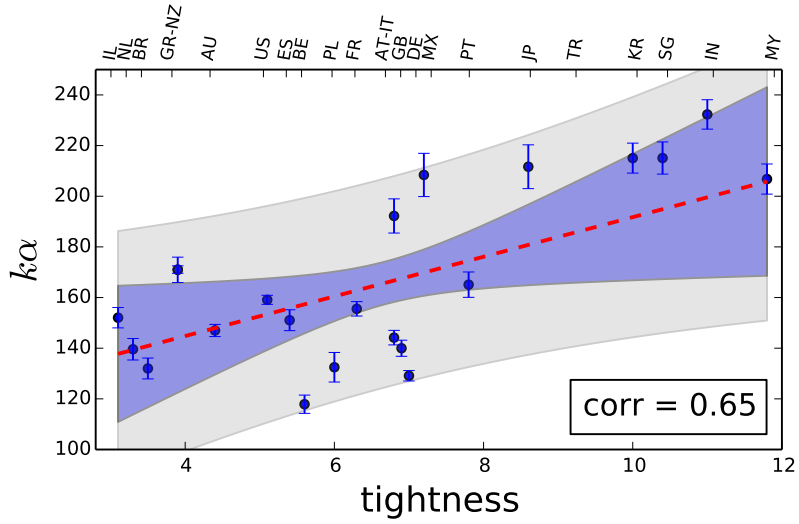


Figura 5.7: Relação entre a pressão social do modelo de agentes e o *tightness score*. As barras de erros foram calculadas a partir de 95% de confiança do *bootstrap*. A linha vermelha representa o melhor ajuste linear ponderado pelo erros dos pontos, onde a região em azul representa o intervalo de de 68% de confiança para essa regressão. A região em cinza é o intervalo de predição de 68% da regressão linear. Observa-se claramente que a pressão social medida pela teoria cresce com o índice de *rigidez/flexibilidade* do país, sendo que as duas medidas apresentam uma correlação de 0.65.

.

são. A região em cinza é o intervalo de predição de 68% da regressão linear. Observa-se claramente que a pressão social medida pela teoria de agentes adaptativos cresce com o índice de *rigidez/flexibilidade* do país, sendo que as duas medidas apresentam uma correlação de 0.65.

Concluímos assim, que a modelagem de agentes e a aproximação de campo médio, apesar de extrema simplificação dos fenômenos sociais, fornecem boas medidas para que se possa entender alguns aspectos de culturas distintas. Além disso, é necessária uma maior investigação para que possamos entender com mais profundidade a validade das hipóteses usadas para o cálculo das proporções das tendencias corroborativas e as pressões sociais de cada país.

País	sigla	*rigidez/flexibilidade*	n_{MFQ}	$k\alpha$
Austrália	AU	4.4	2523	147(2)
Áustria	AT	6.8	147	192(7)
Bélgica	BE	5.6	201	118(4)
Brasil	BR	3.5	464	132(4)
França	FR	6.3	513	156(3)
Alemanha	DE	7.0	1000	129(2)
Grécia	GR	3.9	120	171(5)
Índia	IN	11.	996	232(6)
Israel	IL	3.1	175	152(4)
Itália	IT	6.8	257	144(3)
Japão	JP	8.6	198	212(8)
Coreia do Sul	KR	10.	220	215(6)
Malásia	MY	11.	119	207(6)
México	MX	7.2	339	208(9)
Holanda	NL	3.3	575	140(4)
Nova Zelândia	NZ	3.9	932	171(1)
Polônia	PL	6.0	192	132(6)
Portugal	PT	7.8	188	165(5)
Singapura	SG	10.	347	215(6)
Espanha	ES	5.4	228	151(4)
Turquia	TR	9.2	149	330(36)
Estados Unidos da América	US	5.1	113012	159(2)
Reino Unido	GB	6.9	5306	140(3)

Tabela 5.1: Dentre os 120 países para os quais obtivemos dados do Questionário dos Fundamentos Morais [2], selecionamos 23 que tinham pelo menos 200 respondentes (válidos ou não válidos) e também faziam parte da lista de países analisados no artigo[44]. Estes países, assim como os seus índices de *rigidez/flexibilidade* e pressões sociais vezes o número médio de vizinhos ($k\alpha$) estão listados na tabela ao lado. A coluna n_{MFQ} indica o número de respondentes válidos de cada país; ou seja, o número de respondentes que completaram o Questionário dos Fundamentos Morais.

[6] Conclusão

> "After several unsuccessful attempts to weld my results together into such a whole, I realized that I should never succeed. The best that I could write would never be more than philosophical remarks; my thoughts were soon crippled if I tried to force them on in any single direction against their natural inclination.—And this was, of course, connected with the very nature of the investigation. For this compels us to travel over a wide field of thought criss-cross in every direction.—"
>
> — Ludwig Wittgenstein, 1889-1951

Moralidade, ideologias políticas, influência social são temas de profundos e incontáveis debates filosóficos e científicos. Nesta tese, apresentamos um modelo de agente Bayesianos interagentes que é capaz de mimetizar diversas características do aprendizado moral. Para tanto, foi necessário uma revisão bibliográfica de áreas com pouca mas crescente interseção na pesquisa tradicional em Física, como Psicologia Social, Sociologia, Ciências Políticas além de áreas um pouco mais correlatas como Biologia e Neurociência.

A grande vantagem da modelagem de agentes, e de modelos de Mecânica Estatística em geral, é que ela permite, a partir de uma descrição matemática estilizada do comportamento de indivíduos ou partículas, obter uma descrição do comportamento macroscópico do sistema em estudo.

Um dos principais resultados apresentado nesta tese é que o estilo cognitivo do agente Bayesiano depende da complexidade da interação social na Fase 1, sendo que o estilo cognitivo induz uma associação

estatística a uma filiação política depois que a sociedade de agente atinge o estado estacionário da Fase 2.

A Fase 1 é uma mímica do aprendizado de pessoas durante a pré adolescência, ou da infância até o inicio da fase adulta, já a Fase 2 tenta reproduzir o aprendizado de pessoas na fase adulta. Na Fase 2, a estratégia cognitiva do agente é cristalizada, ou seja, ela deixa de evoluir de acordo com a informação processada, como acontece na Fase 1. No entanto, o agente ainda é capaz de mudar a direção de sua matriz moral, isso faz com que as matrizes morais dos agentes tenham uma assinatura estatística comparável às matrizes morais de pessoas que responderam o questionário sobre Fundamentos Morais. O mais importante resultado de nossa abordagem é que a complexidade vivenciada pelo agente na Fase 1, é positivamente correlacionada com a chance da sociedade formada por esse tipo de agente ter a mesma assinatura estatística de pessoas liberais.

A ideia de usarmos uma descrição puramente probabilística é que esta nos fornece um arcabouço teórico coerente para lidarmos com problemas onde a informação é incompleta[22, 66, 27][1]. Além disso, o algoritmo de aprendizado Bayesiano tem duas importantes características, a primeira é que ele surge de forma natural no contexto probabilístico e pode ser deduzido a partir de princípios de informação mínima[22]. A segunda propriedade é que esse algoritmo aprende de forma *ótima*. O algorítmo de aprendizado usado em nossa modelagem pode ser considerado ótimo pois ele pode ser obtido através da otimização funcional do aprendizado de uma rede neural[2]. Além disso, como foi mostrado por Neirotti e Caticha[3] usando programação evolutiva, algoritmos de classificação submetido a pressões evolutivas por menores erros de generalização são levados a algoritmos parecidos com o Bayesiano tanto em performance quanto em forma funcional. Portanto, se existir uma pressão evolutiva em seres humanos para minimizar o erros de classificação de assuntos durante interações sociais, esperamos que a descrição Bayesiana probabilística do aprendizado seja adequada.

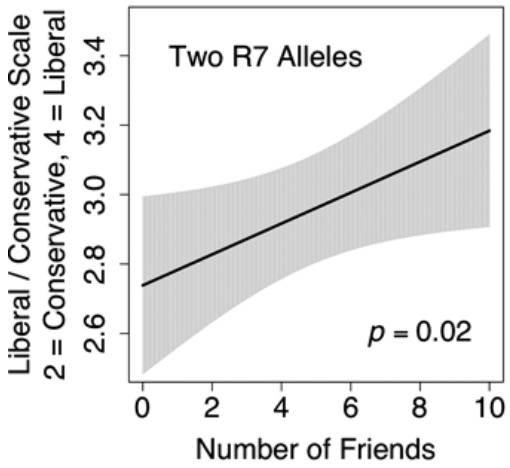
É marcante que alguns resultados de nossa teoria estejam em con-

[1] A. Caticha. Entropic Inference and the Foundations of Physics. *Brazilian Chapter of the International Society for Bayesian Analysis-ISBrA, Sao Paulo, Brazil*, 2012; E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, 2003; and R. T. Cox. Probability, Frequency and Reasonable Expectation. *American Journal of Physics*, 14(1):1, 1946

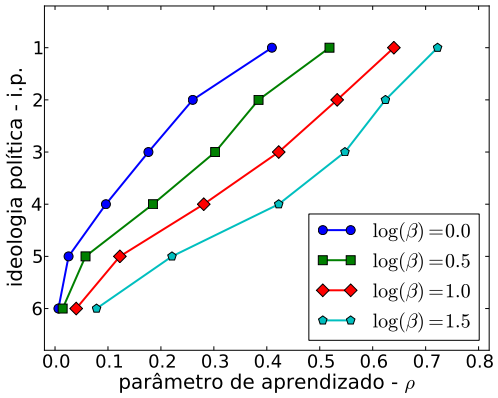
[2] O. Kinouchi and N. Caticha. Optimal generalization in perceptrons. *Journal of Physics A: Mathematical and*, 25:6243–6250, 1992

[3] J. Neirotti and N. Caticha. Dynamics of the evolution of learning algorithms by selection. *Physical Review E*, 67(4):041912, Apr. 2003

formidade com observações experimentais que exploram os motivos do liberalismo em diferentes contextos. Uma importante evidência experimental, no contexto de genética, que conecta complexidade de informação e liberalismo é apresentada por Settle *et al* [102] onde o número de amigos que a pessoa teve na infância está correlacionada com o liberalismo, pelo menos para pessoas que apresentam dois alelos do gene **DRD4-R7**. Os autores desse trabalho apontam cautelosamente que seu estudo não é suficiente para concluir que a presença de um gene especifico é causa da ideologia política, mas sim que existem evidências de que relação entre gene e ambiente influenciam nesse comportamento.



(a) Figura retirada de [29]: O Liberalismo está positivamente correlacionado com o número de amigos na infância de pessoas com duplos alelo do *R7*.



(b) Ideologia Política (pa= 1-Muito liberal, pa=7-Muito Conservador) está correlacionada com ρ, que é uma medida da diversidade moral aprendida durante a Fase 1.

Figura 6.1: Agentes Bayesianos com maior exposição a informação na fase 1 (maior ρ) apresentam assinaturas estatista de opiniões semelhantes a de pessoas liberais, além disso, o liberalismo deve estar positivamente correlacionado com a diversidade de informação obtida na infância.

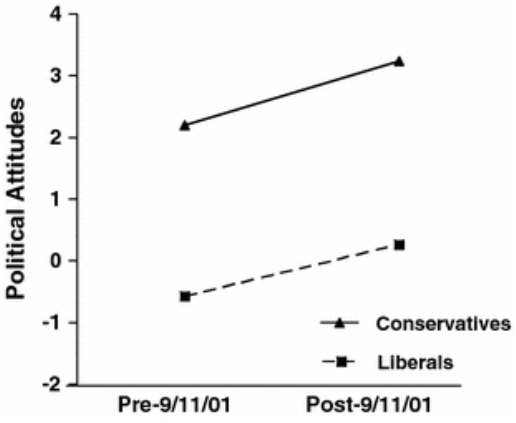
Ainda no contexto de [102] é possível nos perguntarmos qual é a interpretação genética dos resultados de nossa teoria. Nossa teoria não explica os mecanismo genéticos que fazem com que pessoas tenham diferentes estratégias cognitivas, no entanto, podemos levantar hipóteses sobre quais fenômenos poderiam fazer com que as pessoa tivesse acesso a mais complexidade de informação na Fase 1, como maior números de amigos, maior velocidade na aquisição de informação moral, prolongamento do período da adolescência, entre outras.

Além disso, é interessante notar que as fases de aprendizado do

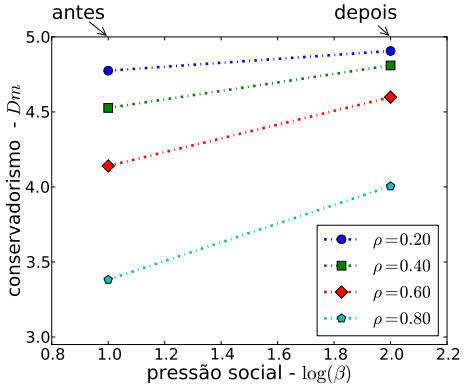
agente Bayesiano conseguem reproduzir qualitativamente os diferentes perfis exploratórios de liberais e conservadores[71]. Observou-se no experimento probabilístico realizado por Fazio *et al.*[4] que pessoas liberais (assim como agentes com ρ próximo de 1) aderem à estratégias mais arriscadas e exploratórias, mas que conferem vantagens no fim do experimento, enquanto pessoas conservadores (assim como agentes com ρ próximos de zero) aderem a estratégias mais prudentes e menos informativas mas que lhes garantem uma vantagem no início do experimento.

[4] N. J. Shook and R. H. Fazio. Political ideology, exploration of novel stimuli, and attitude formation. *Journal of Experimental Social Psychology*, 45(4):995–998, July 2009

O tempo de adaptação não foi usado na formulação teórica do modelo, ele é uma consequência física do processo de troca de informação na sociedade e portanto uma previsão do modelo. Diferentes estilos cognitivos, através da interação social, geram diferentes tempos de adaptação. A existência de uma transição de fase entre sociedades moralmente ordenadas para totalmente desordenadas talvez não deva existir na realidade. No entanto, esperamos que este modelo possa ser aplicado a outros cenários culturais relevantes, onde grupos dos dois lados da transição possam ser encontrados.



(a) Figura retirada de [83]. Índice de atitude política média medida antes e depois dos ataque terrorista de 11/09/2001. Quanto maior o índice de atitude política maior a ideologia conservadora.



(b) O número efetivo de dimensões morais para dois valores de pressão social, antes e depois de uma ameaça. Se ameaças acarretam em aumento da pressão social, a assinatura estatística de liberais se torna mais parecida com a de conservadores.

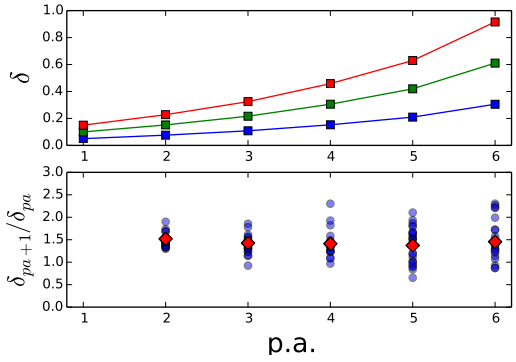
Figura 6.2: sob o aumento do parâmetro de pressão social β a sociedade tenderá estatisticamente a ficar mais conservadora, como é mostrado nas figuras 4.5 e 4.6

Outra predição do modelo é que sob o aumento do parâmetro de pressão social β a sociedade tenderá estatisticamente a ficar mais con-

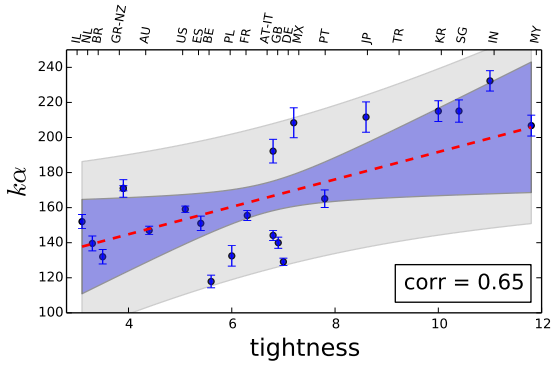
servadora, como é mostrado na figura 6.2. Esse efeito de crescimento da pressão social pode estar por trás de resultados como os de Bonano *et al* [15] e de Nail *et al* [83] sobre o crescimento do conservadorismo nos Estados Unidos após o ataque de 11/09. No entanto, Nail *et al* [84] mostraram que mesmo que não exista uma intervenção social direta em forma de ameaça externa para que a sociedade se torne mais conservadora, um efeito equivalente pode ser obtido, pelo menos momentaneamente, ao se fazer com que o indivíduo se imagine numa situação de ameaça ou conforto. Isso sugere que nossa interpretação do parâmetro β como pressão social pode ser entendida como um parâmetro que pode ser auto regulado dinamicamente através da extração da informação do contexto social.

No capítulo 5, usando os dados de matrizes morais de pessoas de 23 países, evidenciamos que a modelagem de agentes proposta nesse trabalho e também no trabalho de Caticha e Vicente[5] são suficientemente abrangentes para descrever estatisticamente os dados de opinião moral de pessoas de diferentes ideologias políticas e culturas distintas.

[5] N. Caticha and R. Vicente. Agent-Based Social Psychology: From Neurocognitive Processes To Social Data. *Advances in Complex Systems*, 14(05):711, 2011



(a) Na figura acima é apresentada a tendência corroborativa em função da filiação política. Na figura abaixo os pontos azuis representam proporção das tendências corroborativas entre pessoas de ideologias políticas consecutivas, na escala liberal conservador, para cada país. Os pontos vermelhos são as medianas das proporções das tendências corroborativas.



(b) Relação entre a pressão social do modelo de agentes e o *tightness score*. Observa-se que a pressão social medida pela teoria cresce com o índice de *rigidez/flexibilidade* do país, sendo que as duas medidas apresentam uma correlação de 0.65.

Para que conseguíssemos analisar os dados multiculturais sob a perspectiva de nosso modelo fizemos duas importantes hipóteses. Pri-

meiramente assumimos que a ideologia política de indivíduos é função exclusiva de sua tendência corroborativa. E segundo, que a pressão social percebida pelos indivíduos de uma cultura é a mesma para todos os indivíduos dessa cultura não dependendo da sua ideologia política.

Um importante resultado obtido a partir dessa análise é que razão entre as tendencias corroborativas de indivíduos com filiações políticas consecutivas na escala liberal conservador é próximo de uma constante ($\delta_{pa+1}/\delta_{pa} \approx 1.5$). De fato, ainda não conseguimos obter uma justificativa ou medida empírica relacionada com esse resultado, no entanto, podemos especular que, no contexto do modelo de agentes Bayesianos, a ideologia política auto-declarada do indivíduo está relacionada com a dependência não-linear do parâmetro ρ com a quantidade de informação obtida pelo agente na primeira fase de aprendizado.

Outro resultado marcante obtido foi a alta correlação entre a medida de pressão social dos diversos países e a medida de *rigidez/flexibilidade* cultural proposta por Gelfand *et al.* [44]. Como discutido nesse trabalho, países rígidos, assim como o esperado por nossa teoria, são países com normas sociais fortes e pouca tolerância para comportamentos desviantes. Além disso, o grau de rigidez de um pais está relacionado com um sistema complexo e integrado de ameaças a sociedade de origem tanto ecológica quanto humana.

Uma característica importante de nossa teoria é que ela é livre de semântica. Ou seja, a matemática usada para definir os vetores morais não é capaz de diferenciar se uma componente se refere à fundação moral de justiça ou lealdade. Acreditamos que esse importante aspecto deva ser investigado sob uma perspectiva evolutiva, de forma que se possa compreender a emergência das diferentes dimensões morais e construir um arcabouço matemático para a incorporação da semântica na teoria. Além disso, essa característica de nosso modelo abre a oportunidade de que ele seja útil para modelar outros problemas que não o do aprendizado moral.

[A] Redes Neurais

> One of the great intellectual challenges for the next few decades is the question of brain organization. What is the basic mechanism for storage of memory? What are the processes that serve as the interface between the basically chemical processes of the body and the very specific and non-statistical operations of the brain? Above all, how is concept formation achieved in the human brain? I wonder whether the spirit of the physics that will be involved in these studies will not be akin to that which moved the founders of the "rational foundation of thermodynamics".
>
> —C.N. Yang, 1922-

O neurônio, de forma extremamente simplificada, pode ter seu funcionamento descrito como uma série de impulsos elétricos que são recebidos pelas partes de sua célula chamadas de dendritos. Por eles, os impulsos são enviados para o corpo da célula, sendo que durante o envio os impulsos podem ser atenuados ou amplificados, a depender do dendrito. A combinação desses impulsos é recebida pelo corpo celular, que modifica esses sinais de alguma maneira e os envia na forma de pulsos através do axônio. Estes pulsos se propagam até as terminações nervosas do axônio, de onde são transmitidos para dendritos de outros neurônios.

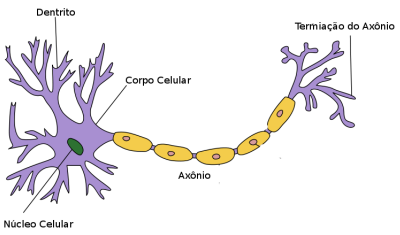


Figura A.1: Ilustração de um neurônio retirada de [3]

É claro que biologicamente esse processo é muito mais complexo, sendo esses impulsos causados por trocas de íons que têm suas taxas controladas por vários fatores. No entanto, esse esquema de funcionamento do neurônio serve de inspiração para o modelo matemático de rede neural que iremos usar. Dentre as várias descrições matemáticas

do neurônio, a mais simples é a descrição do **perceptron**, introduzida por Rosenblatt no final da década de 1950 [98]. Por essa descrição, representaremos os impulsos que chegam ao axônio como um vetor de $N-$dimensões $\mathbf{X} = (X_1, X_2, \cdots, X_N)$. Esse impulso de entrada é ponderado por um vetor de pesos sinápticos $\mathbf{J} = (J_1, J_2, \cdots, J_N)$ de mesma dimensão. O impulso de saída é modelado como uma função desses dois vetores, $f(\mathbf{X}, \mathbf{J})$.

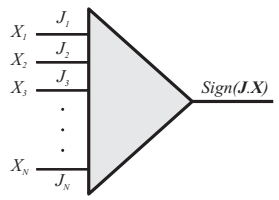


Figura A.2: Representação matemática de uma rede neural

Faremos ainda uma descrição mais simplista do neurônio [34], fazendo com que o perceptron seja basicamente um classificador booleano do vetor de entrada $\mathbf{X} \in \mathbb{R}^N$, onde $f(\mathbf{J} \cdot \mathbf{X}) = \sigma[\mathbf{J}, \mathbf{X}] = Sign(\mathbf{J}\mathbf{X})$. Assim, o perceptron separa os vetores de entrada em dois hiperplanos, um que está na direção do perceptron e outro na direção contrária.

[A.1] Perceptron Booleano

O modelo de rede neural do perceptron, apesar de simples, é muito útil para lidar com problemas em que existe uma classificação linear de dados. Dentre os vários tipos de algoritmos de aprendizado para redes neurais, um dos mais conhecidos são os algoritmos de aprendizado supervisionado do perceptron. Nesses algoritmos, vetores de entrada (que chamaremos de assuntos), provindos independente de uma distribuição de probabilidade $P_0(\mathbf{X})$, são apresentados juntamente com uma classificação pré-estabelecida $(\mathbf{X}_\mu, \sigma_\mu)$, sendo $\mu = 1, \ldots, p$. O algoritmo de aprendizado fará com que o perceptron, usando essa informação, "mude seus pesos sinápticos" $\mathbf{J}$ a fim de que sua classificação coincida com a classificação pré-estabelecida dos assuntos, σ_μ.

Existem vários paradigmas de aprendizado para o perceptron [1], dentre eles podemos destacar o aprendizado on-line e aprendizado off-line. No aprendizado off-line a direção do perceptron é encontrada usando todos os exemplos de uma vez. Neste texto, nos restringiremos apenas na discussão dos algoritmos de aprendizado on-line, já que esse pode ser interpretado mais facilmente sobre a perspectiva de aprendizado por reforço. Nele, um perceptron $\mathbf{J}$, ao qual chamaremos de **estudante** muda a direção de seus pesos sinápticos de acordo com

[1] A. Engel and C. Van den Broeck. *Statistical mechanics of learning*. Cambridge University Press, Cambridge, 2004

a equação (A.1)

$$\mathbf{J}_{\mu+1} = \mathbf{J}_\mu - \frac{1}{N}\nabla_{\mathbf{J}} V_\mu, \tag{A.1}$$

onde $\nabla_{\mathbf{J}} V_\mu \equiv -W_\mu \sigma_B^\mu \mathbf{X}_\mu$ por definição. O índice $\mu = 1, \ldots, p$ indica o tempo de treinamento do algoritmo. A classificação *a priori* do assunto é provida por um outro perceptron **B**, chamado de **professor**, através do sinal do produto escalar com o professor, $\sigma_B^\mu = sign\left(\mathbf{B}.\mathbf{X}_\mu\right)$.

Com isso, vemos que os pesos sinápticos do perceptron aluno mudam na direção $\sigma_B^\mu \mathbf{X}$, conhecido como termo Hebbiano. O fator de proporcionalidade W_μ é chamado de função de modulação. É esperado que, após um tempo de treinamento suficientemente longo, o aluno classifique os assuntos

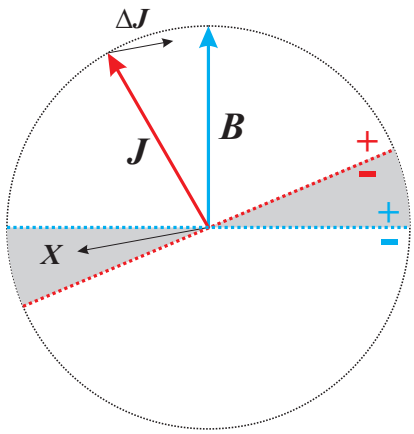


Figura A.3: *Representação do hiperplano do perceptron e do algoritmo de aprendizado.* O vetor exemplo **X** é classificado pelos perceptrons professor **B** e aluno **J**, em $+1$ ou -1, dependendo de que lado do hiperplano de cada um o vetor se encontra. Na região escura os vetores de exemplo são classificados diferentemente por professor e aluno. O vetor $\Delta\mathbf{J}$ é a correção (proporcional ao termo Hebbiano $\sigma_B\mathbf{X}$) na direção do aluno.

Existem vários tipos de função de regulação W_μ estudadas na literatura, quando $W = 1$ temos o algoritmo de aprendizado Hebbiano. O algoritmo do perceptron propriamente dito é tido quando $W_\mu = \Theta\left(-h_\mu \sigma_{\mathbf{B}}\right)$, sendo $\Theta(x)$ a função degrau de Heaviside[2], onde

$$h = \frac{\mathbf{J}\cdot\mathbf{X}^\mu}{|\mathbf{J}|} \qquad \text{e} \qquad b = \mathbf{B}\cdot\mathbf{X}^\mu. \tag{A.3}$$

Diferentemente do algoritmo de aprendizado hebbiano, que sempre muda o vetor de pesos sinápticos do estudante, o algoritmo do perceptron estudante só muda os pesos sinápticos quando ele discorda do professor. O potencial de aprendizado do algoritmo de Hebb é $V_\mu = -h_\mu \sigma_B$ e o potencial do algoritmo do perceptron é $V_\mu =$

[2] A função de Heaviside é dada por,

$$\Theta(x) = \begin{cases} 1 & se \quad x \geq 0, \\ 0 & se \quad x < 0. \end{cases} \tag{A.2}$$

$-h_\mu \sigma_B \Theta\left(-h_\mu \sigma_B\right)$.

A medida natural de quanto o aluno está alinhado com o professor é conhecida como *overlap* ρ que é simplesmente o produto escalar entre os vetores sinápticos do professor e do aluno dividido por suas normas.

$$\rho = \frac{\mathbf{B} \cdot \mathbf{J}}{|\mathbf{B}|\,|\mathbf{J}|}. \tag{A.4}$$

Na próxima seção veremos que a variação de ρ em relação ao tempo de treinamento está diretamente relacionada com a eficiência do algoritmo de aprendizado para diferentes funções de modulação.

[A.2] Aprendizado Ótimo

Uma ferramenta útil para medirmos a convergência do algoritmo aprendizado do perceptron é o **erro de generalização** [71, 117][3] que é definido como a média, sobre a distribuição de assuntos, dos erros instantâneos

$$e_P = \frac{1}{2}\left(1 - \sigma_J \sigma_B\right) = \frac{1}{2}\Theta\left(-hb\right).$$

Em outras palavras, o erro de generalização é a probabilidade de que o estudante classifique um assunto aleatório da mesma maneira que o professor,

$$\begin{aligned} e_G &= \int \mathrm{d}\mathbf{X} P_0\left(\mathbf{X}\right) e_P\left[\sigma_J\left(\mathbf{X}\right)\sigma_B\left(\mathbf{X}\right)\right], \\ &= \int \mathrm{d}\mathbf{X} P_0\left(\mathbf{X}\right) \frac{1}{2}\Theta\left[-\left(\mathbf{J}\cdot\mathbf{X}\right)\left(\mathbf{B}\cdot\mathbf{X}\right)\right]. \end{aligned} \tag{A.5}$$

No *Limite Termodinâmico*, que acontece quando $N \to \infty$, podemos provar que o erro de generalização depende exclusivamente do *overlap* entre o estudante e o aluno ρ, de acordo com a seguinte expressão, [4].

$$\begin{aligned} e_G &= \int \mathrm{d}h\mathrm{d}b P\left(h,b\right)\Theta\left(-hb\right) \\ &= \frac{1}{\pi}\cos^{-1}\left(\rho\right) \end{aligned} \tag{A.6}$$

Nesse limite, pode-se provar, usando o *Teorema Central do Limite*, que a distribuição $P\left(h,b\right)$ é uma distribuição normal de média nula e matriz

[3] O. Kinouchi and N. Caticha. Optimal generalization in perceptrons. *Journal of Physics A: Mathematical and*, 25:6243–6250, 1992; and R. Vicente, O. Kinouchi, and N. Caticha. Statistical mechanics of online learning of drifting concepts: a variational approach. *Machine learning*, 201:179–201, 1998

[4] A. Engel and C. Van den Broeck. *Statistical mechanics of learning*. Cambridge University Press, Cambridge, 2004

de correlação dada por

$$C = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \tag{A.7}$$

onde a correlação ρ entre o estudante e professor é exatamente *overlap* definido em (A.4). De forma explícita, a distribuição de probabilidade dos assuntos em relação ao perceptron estudante e professor é

$$P(h,b) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp -\frac{h^2 - 2\rho hb + b^2}{2(1-\rho^2)}. \tag{A.8}$$

O nosso objetivo é encontrar uma função de modulação W_μ que faça com que o perceptron aprenda com o menor número de exemplos possíveis. Introduzindo as variáveis, $J_\mu = |\mathbf{J}_\mu|$ e $R_\mu = \mathbf{J}_\mu \cdot \mathbf{B}$ e lembrando que a equação de aprendizado é dada por,

$$\mathbf{J}_{\mu+1} = \mathbf{J}_\mu + \frac{1}{N} W_\mu \sigma_B \mathbf{X}_\mu, \tag{A.9}$$

segue que a dinâmica dessas variáveis é descrita por,

$$R_{\mu+1} = R_\mu + \frac{1}{N} W_\mu \sigma_B^\mu b_\mu; \tag{A.10}$$

$$\begin{aligned} J_{\mu+1} &= J_\mu \left[1 + \frac{1}{N} \left(2\frac{W_\mu}{J_\mu} \sigma_B^\mu h_\mu + \frac{W_\mu^2}{J_\mu^2} \right) \right]^{1/2}; \\ &\approx J_\mu \left[1 + \frac{1}{N} \left(\frac{W_\mu}{J_\mu} \sigma_B^\mu h_\mu + \frac{W_\mu^2}{2J_\mu^2} \right) \right]; \end{aligned} \tag{A.11}$$

onde consideramos $|X_\mu| \propto \mathcal{O}(N)$ e desprezamos termos de ordem superior á $1/N$. Com isso, podemos escrever a evolução do *overlap* entre aluno e professor como,

$$\begin{aligned} \rho_{\mu+1} &= \frac{R_{\mu+1}}{J_{\mu+1}}, \\ &= \frac{R_\mu + W_\mu \sigma_B^\mu b_\mu}{J_\mu \left[1 + \frac{1}{N} \left(2\frac{W_\mu}{J_\mu} \sigma_B^\mu h_\mu + \frac{W_\mu^2}{J_\mu^2} \right) \right]^{1/2}}; \\ &\approx \rho_\mu + \frac{1}{N} \left[(b_\mu - \rho_\mu h_\mu) \sigma_B^\mu \frac{W_\mu}{J_\mu} - \frac{\rho_\mu W_\mu^2}{2J_\mu^2} \right]. \end{aligned} \tag{A.12}$$

As equações (A.10) e (A.11) são estocásticas, elas podem ser escritas

como equações diferenciais no Limite Termodinâmico ao definirmos $\alpha = \mu/N$, de onde teremos $\mathrm{d}\alpha = 1/N$,

$$\frac{\mathrm{d}R}{\mathrm{d}\alpha} = \langle W\sigma_B b\rangle\,, \tag{A.13}$$

$$\frac{\mathrm{d}J}{\mathrm{d}\alpha} = J\left\langle \frac{W}{J}\sigma_B h + \frac{W^2}{2J^2}\right\rangle; \tag{A.14}$$

$$\frac{\mathrm{d}\rho}{\mathrm{d}\alpha} = \left\langle (b-\rho h)\,\sigma_B\frac{W}{J} - \rho\frac{W^2}{2J^2}\right\rangle; \tag{A.15}$$

onde $\langle\cdots\rangle = \int \mathrm{d}\mathbf{X}P_0(\mathbf{X})\,[\cdots]$ é a média tomada sobre a distribuição de exemplos.

Para maximizarmos a taxa de aprendizado do algoritmo devemos minimizar a taxa do erro de generalização. A ideia do algoritmo de aprendizado ótimo[5], vem do fato que o erro de generalização é função exclusiva do *overlap* entre o professor e o aluno, e esse é um funcional da função de modulação, de onde segue,

[5] O. Kinouchi and N. Caticha. Optimal generalization in perceptrons. *Journal of Physics A: Mathematical and*, 25:6243–6250, 1992

$$\frac{\mathrm{d}e_G}{\mathrm{d}\alpha}[W] = \frac{\mathrm{d}e_G}{\mathrm{d}\rho}\frac{\mathrm{d}\rho}{\mathrm{d}\alpha}[W]$$

Cada função de modulação W leva a uma dinâmica de aprendizado diferente, então, podemos fazer uma derivada funcional de *overlap* em relação à função de modulação para encontrarmos uma função que minimiza o tempo de aprendizado,

$$\frac{\delta}{\delta W}\left[\frac{\mathrm{d}\rho}{\mathrm{d}\alpha}\right] = 0. \tag{A.16}$$

Portanto, a partir de (A.15) e (A.16) encontramos que a função que minimiza o tempo de aprendizado é dada por

$$W^*_\mu = J_\mu\left(\kappa_\mu - z_\mu\right), \tag{A.17}$$

com

$$\kappa_\mu = \frac{\sigma^\mu_B b_\mu}{\rho_\mu} \qquad \text{e} \qquad z_\mu = \sigma^\mu_B h_\mu. \tag{A.18}$$

Percebemos que a função de aprendizado ótima escrita na forma (A.17) depende de dois tipo de variáveis:

Visiveis: Que são as variáveis que o estudantes tem acesso,

$$\mathcal{V} = \{h_\mu, \sigma_B^\mu\}$$

Invisiveis: Que são as variáveis que o estudante não tem acesso,

$$\mathcal{H} = \{|b_\mu|\}$$

O algoritmo de aprendizado ótimo será realista somente se a função de modulação W^* for calculada tomando-se a média sobre as variáveis invisíveis dado o conhecimento das visíveis. Com isso, a função de modulação ótima será dada por,

$$\bar{W} = \langle W^* \rangle_{\mathcal{H}|\mathcal{V}} \,, \tag{A.19}$$

onde $\langle \cdots \rangle_{\mathcal{H}|\mathcal{V}} = \int \mathrm{d}\mathcal{H} P(\mathcal{H}|\mathcal{V}) [\cdots]$.

Sabendo que a distribuição de probabilidade conjunta das variáveis visíveis e invisíveis é dada por $P(\mathcal{H}, \mathcal{V}) = P(\mathcal{H}) P(\mathcal{H}|\mathcal{V})$ e lembrando da regra de Bayes,

$$P(\mathcal{H}|\mathcal{V}) = \frac{P(\mathcal{H}, \mathcal{V})}{\int \mathrm{d}\mathcal{H} P(\mathcal{H}, \mathcal{V})}, \tag{A.20}$$

podemos ver que a função de modulação ótima deve ser calculada pela integral

$$\bar{W} = \frac{\int \mathrm{d}|b| P(b,h) W^*}{\int \mathrm{d}|b| P(b,h)}. \tag{A.21}$$

Essa integral é calculada mais facilmente vendo que

$$\int \mathrm{d}|b| [\cdots] = \int \mathrm{d}b [\Theta(b) - \Theta(-b)] [\cdots].$$

Logo a expressão para função de modulação do aprendizado ótimo é

$$W(\rho_\mu, J_\mu, z_\mu) = \frac{1}{\sqrt{2\pi}} J_\mu \lambda_\mu \exp\left(-\frac{z_\mu^2}{2\lambda_\mu^2}\right) \frac{1}{H(-z_\mu/\lambda_\mu)}, \tag{A.22}$$

onde

$$\lambda = \tan(\pi e_g) = \frac{\sqrt{1-\rho^2}}{\rho} \tag{A.23}$$

e

$$H(x) = \frac{1}{2\pi}\int_x^\infty dy \exp(\frac{-y^2}{2}) = \frac{1}{2}\text{ercf}\left(\frac{x}{\sqrt{2}}\right) \tag{A.24}$$

Na figura 3.1 da sessão 3.1.1 (repetida abaixo), representamos a curva da função de aprendizado para alguns valores de ρ com o tamanho do vetor aluno fixo em $J = 1$ em função de z. Podemos perceber, que quanto mais o vetor aluno aprende, ou seja, quanto maior o valor de ρ, mais o aluno aprende com os exemplos que ele classifica de maneira diferente do professor ($z < 0$) em comparação com os exemplos que ele classifica do mesmo modo ($z > 0$).

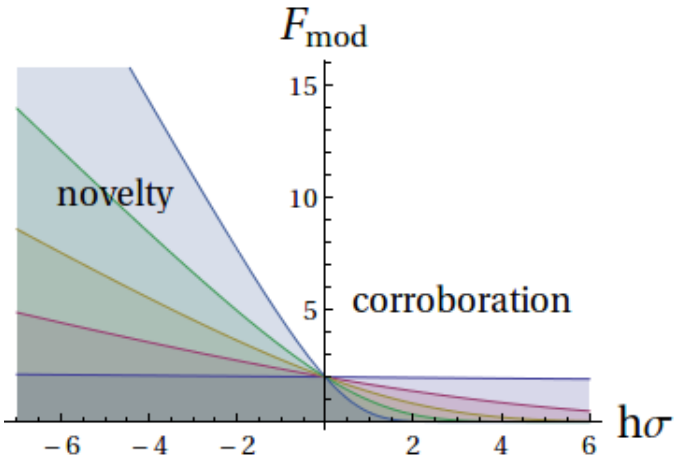


Figura A.4: Dependência da função de modulação com a quantidade de troca de informação medido através de ρ. Quanto mais o aluno aprende, maior é a amplitude da função de modulação para informações que tem novidade, $h\sigma < 0$; e menor é a amplitude para informações que não tem novidade, $h\sigma > 0$.

⌈B⌉ Aprendizado Bayesiano

> Probability theory is nothing but common sense reduced to calculation.
>
> — Pierre-Simon Laplace,1749–1827

⌈B.1⌉ Probabilidade

Ainda hoje, o significado ou interpretação da probabilidade é um tema de grande debate acadêmico[22][1]. A interpretação mais comum é a *frequentista* onde a probabilidade de um evento deve ser calculada de maneira experimental através da razão do número de vezes que o evento foi observado pelo número total de tentativas ou experimentos. Essa interpretação tem um grande apelo e é considerada por muitos como uma interpretação *objetiva* pois a probabilidade de um evento ocorrer nada tem haver com o observador. Apesar de a primeira vista essa interpretação ter um grande apelo científico, ela não é totalmente livre de controvérsias pois deixa algumas dúvidas a respeito da natureza da aleatoriedade. Por exemplo, é possível calcular a probabilidade de eventos não reprodutíveis como a probabilidade de existir vida em Marte ou mesmo de fazer sol amanhã? Além disso, qual é o grau de precisão necessário para se dizer que um evento é reprodutível[2]?

Já a interpretação *Bayesiana* diz que a probabilidade deve representar o grau de certeza ou a crença de que uma proposição[3] é verdadeira. Apesar dessa interpretação ter se tornado mais popular somente nas últimas décadas ela remonta a Bernoulli e Laplace. De fato, seguindo essa tradição de pensamento, o físico americano Richard T. Cox mos-

[1] A. Caticha. Entropic Inference and the Foundations of Physics. *Brazilian Chapter of the International Society for Bayesian Analysis-ISBrA, Sao Paulo, Brazil*, 2012

[2] Podemos nos perguntar agora o quanto a interpretação frequentista da probabilidade é objetiva.

[3] Um exemplo de proposição é: O evento A ocorreu.

trou em 1946 que admitindo alguns *critérios de coerência* a teoria da probabilidade é a maneira adequada para se tratar de grau de incertezas de proposições[27, 28][4]. De fato, os critérios de coerência de Cox nada mais são que um conjunto de axiomas usados para formular a teoria de probabilidade e podem ser sintetizados através da seguinte frase[5][27],

> *"...quanto menos verossímil é um evento ocorrer mais verossímil é ele não ocorrer. A ocorrência dos dois eventos não será mais verossímil e geralmente será menos verossímil do que o menos verossímil dos dois. No entanto, a ocorrência de pelo menos um dos eventos não é menos verossímil e é geralmente mais verossímil do que a ocorrência de ambos."*

[4] R. T. Cox. Probability, Frequency and Reasonable Expectation. *American Journal of Physics*, 14(1):1, 1946; and R. T. Cox. The Algebra of Probable Inference. *American Journal of Physics*, 31(1):66, 1963

[5] Tradução livre de:*"...the less likely is an event to occur the more likely it is not to occur. The occurrence of both of two events will not be more likely and will generally be less likely than the occurrence of the less likely of the two. But the occurrence of at least one of the events is not less likely and is generally more likely than the occurrence of either."*

Como já é comum em ciência, sempre é possível questionar a validade ou a generalidade de axiomas ou regras, e mesmo as regras de consistências podem ser questionadas ou generalizadas sob certas condições como é evidenciado no artigo de Patriota[6]. No entanto, é inegável a grande generalidade e a aplicabilidade dos conceitos de probabilidade, ainda mais por sua conexão clara com medidas experimentais devido sua correspondência com a frequência de eventos.

[6] A. G. Patriota. A classical measure of evidence for general null hypotheses. *Fuzzy Sets and Systems*, pages 1–15, Mar. 2013

Podemos nos perguntar agora, como podemos inferir a probabilidade de uma asserção? Como mostrado por Ariel Caticha em[22] uma maneira de se inferir a probabilidade de um evento repeitando as regras de coerência de Cox é através maximização da entropia relativa do sistema (também conhecida como divergência de Kullback-Leibler). O trabalho de Ariel Caticha pode ser visto como uma continuação do trabalho do físico E.T. Jaynes[7] que mostra uma conexão mais profunda entre o problema de inferência probabilística, Mecânica Estatística e Teoria da Informação. Um dos resultados interessantes obtido por Ariel Caticha é que a inferência Bayesiana que será descrita na próxima seção é um caso particular da inferência entrópica proposta por ele e Jaynes.

[7] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, 2003

[B.2] Aprendizado Bayesiano Online

Dado um conjunto de variáveis aleatórias $D_t = (y_0, \ldots, y_{t-1})$ e supondo que elas são amostradas de forma independente de acordo com a seguinte distribuição de probabilidade ou *verosimilhança*,

$$P(D_t|\theta) = \prod_{i=0}^{t-1} P(y_i|\theta), \tag{B.1}$$

onde θ é um vetor de parâmetros desconhecidos que queremos estimar. Pela interpretação Bayesiana de probabilidade, é possível assumir uma distribuição probabilidade para eventos onde existe incerteza. Pelo teorema de Bayes temos que a densidade de probabilidade do parâmetro $\boldsymbol{\theta}$ dado o conhecimento das amostras é

$$P(\theta|D_t) = \frac{P(\theta)P(D_t|\theta)}{\int \mathrm{d}\theta' P(\theta')P(D_t|\theta')}. \tag{B.2}$$

A probabilidade $P(\theta|D_t)$ é chamada de *posteriori*, já a quantidade $P(\theta)$ é chamada de *priori* representa o conhecimento prévio sobre os parâmetros. Existe uma extensa literatura[8] de métodos de estatística dedicados à resolução e análise da equaçãoB.2. Neste trabalho nos restringiremos ao algoritmo de aprendizado Bayesiano online proposto por Manfred Opper[88][9] e que também é discutido de forma mais didática em[90, 108].

[8] M. H. DeGroot. *Probability and statistics*. Addison-Wesley, 1989

[9] M. Opper. On-line versus Off-line Learning from Random Examples: General Results. *Physical review letters*, 77(22):4671–4674, Nov. 1996

O aprendizado Bayesiano online é obtido a pela mudança da distribuição a *posteriori* quando um novo exemplo y_t é apresentado. Desta forma, podemos mostrar que a nova distribuição *posteriori* será o produto normalizado entre a antiga *posteriori* e a *verosimilhança* do novo exemplo,

$$P(\theta|D_{t+1}) = \frac{P(\theta|D_t)P(y_t|\theta)}{\int \mathrm{d}\theta' P(\theta'|D_t)P(y_t|\theta')}. \tag{B.3}$$

Quando a quantidade de dados diponível é muito grande a resolução da equação B.3 pode ser proibitiva, para contornar esse problema a *posteriori* $P(\theta|D_t)$ exata pode ser aproximada por uma distribuição paramétrica $P(\theta|A_t)$. Com isso, o aprendizado online é obtido a partir

de dois passos:

- **Processa nova informação**: A partir da apresentação de um novo exemplo a *posteriori* é atualizada através da regra de Bayes

$$P(\theta|A_t, y_t) = \frac{P(\theta|A_t)P(y_t|\theta)}{\int \mathrm{d}\theta' P(\theta'|A_t)P(y_t|\theta')}. \tag{B.4}$$

- **Projeção no Espaço Paramétrico**: Quando a atualização de Bayes é feita, a nova *posteriori* não necessariamente pertence ao espaço paramétrico da distribuição *priore* usada. Para retornarmos ao espaço paramétrico desejado é feito uma projeção da nova *posteriori* nesse espaço através da minimização da divergência de Kullback-Leibler entre as duas distribuições

$$P(\theta|A_t, y_t) \xrightarrow{\mathbf{KL}} P(\theta|A_{t+1}), \tag{B.5}$$

  sendo a divergência de Kullback-Leibler entre as duas distribuições é definida por,

$$\begin{aligned} &KL[P(\theta|A_t, y_t)||P(\theta|A_{t+1})] \\ &\quad = \int \mathrm{d}\theta P(\theta|A_t, y_t) \log \frac{P(\theta|A_t, y_t)}{P(\theta|A_{t+1})}. \end{aligned} \tag{B.6}$$

⌈B.2.1⌉ Ansatz Gaussiano

Considerando um espaço paramétrico gaussiano,

$$\begin{aligned} P(\theta|A_t) &= P(\boldsymbol{\omega}|A_t) \\ &= \frac{1}{|2\pi \boldsymbol{C}_t|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left(\boldsymbol{\omega} - \overline{\boldsymbol{\omega}}_t\right)^T \boldsymbol{C}_t^{-1}\left(\boldsymbol{\omega} - \overline{\boldsymbol{\omega}}_t\right)\right), \end{aligned} \tag{B.7}$$

a minimização de divergência de Kullback-Leibler B.6 equivale a igualarmos os momentos de ambas as distribuições, como é demostramos na secçãoB.4. Assim segue,

$$\begin{aligned} \overline{\boldsymbol{\omega}}_{t+1} &= \frac{\int \mathrm{d}\boldsymbol{\omega}\boldsymbol{\omega} P(\boldsymbol{\omega}|A_t)P(y_t|\boldsymbol{\omega})}{\int \mathrm{d}\boldsymbol{\omega} P(\boldsymbol{\omega}|A_t)P(y_t|\boldsymbol{\omega})}, \\ \boldsymbol{C}_{t+1} &= \frac{\int \mathrm{d}\boldsymbol{\omega}\boldsymbol{\omega}\boldsymbol{\omega}^T P(\boldsymbol{\omega}|A_t)P(y_t|\boldsymbol{\omega})}{\int \mathrm{d}\boldsymbol{\omega} P(\boldsymbol{\omega}|A_t)P(y_t|\boldsymbol{\omega})} - \overline{\boldsymbol{\omega}}_{t+1}\overline{\boldsymbol{\omega}}_{t+1}^T. \end{aligned} \tag{B.8}$$

Fazendo a mudança de variáveis, $\boldsymbol{\omega} = \boldsymbol{u} + \overline{\boldsymbol{\omega}}_t$ e usando a propriedade de densidades gaussianas com média nula, temos,

$$E(xf(x)) = E(x^2)E(f'(x)),$$

e a propriedade, $\frac{\mathrm{d}f(x+y)}{\mathrm{d}x} = \frac{\mathrm{d}f(x+y)}{\mathrm{d}y}$. Podemos mostrar que a média e a covariância serão atualizadas de acordo com as expressões,

$$\begin{aligned} \overline{\boldsymbol{\omega}}_{t+1} &= \overline{\boldsymbol{\omega}}_t - \boldsymbol{C}_t \frac{\partial V_t}{\partial \overline{\boldsymbol{\omega}}_t} \\ \boldsymbol{C}_{t+1} &= \boldsymbol{C}_t - \boldsymbol{C}_t \frac{\partial^2 V_t}{\partial \overline{\boldsymbol{\omega}}_t \partial \overline{\boldsymbol{\omega}}_t^T} \boldsymbol{C}_t \end{aligned} \tag{B.9}$$

onde a função $V_t = -\log \mathcal{E}_t$ pode ser interpretada como uma energia de aprendizado, que depende da função $\mathcal{E}_t$ que é conhecida como evidência e é definida por,

$$\begin{aligned} \mathcal{E}_t &= -E_{\boldsymbol{u}} \left[P(y_t | \overline{\boldsymbol{\omega}}_t + \boldsymbol{u}) \right], \\ &= \int \mathrm{d}\boldsymbol{u} P(\boldsymbol{u}|A_t) P(y_t | \overline{\boldsymbol{\omega}}_t + \boldsymbol{u}). \end{aligned} \tag{B.10}$$

⌈B.3⌉ Perceptron Booleano com Ruído Aditivo e Multiplicativo

Como vimos no apêndice A, no caso do perceptron booleano os dados são sorteados e são escritos na forma $y_t = (\tau_t, \boldsymbol{x}_t)$ onde τ_t é um rotulo, que pode ser binário, ou um números real, e $\boldsymbol{x}_t = (x_t^1, \ldots, x_t^N)$ é um vetor N-dimensional. Considerando que os vetores exemplos são escolhidos independentemente do parâmetro $\boldsymbol{\omega}$ teremos que $P(y|\boldsymbol{\omega}) = P(\tau|\boldsymbol{\omega}, \boldsymbol{x})P(\boldsymbol{x})$. Com isso, as equações (B.8), (B.9) e (B.10) permaneceram inalteradas exceto que $P(y|\boldsymbol{\omega})$ será substituído por $P(\tau|\boldsymbol{\omega}, \boldsymbol{x})$.

Podemos imaginar que existam dois tipos de erros que podem corromper a classificação do exemplo, um ruído multiplicativo e outro aditivo. O ruído multiplicativo inverte a classificação τ do exemplo $\boldsymbol{x}$ com uma probabilidade conhecida ϵ de forma que

$$p(\tau|\boldsymbol{\omega}, \boldsymbol{x}, \sigma_B) = \epsilon \delta_{(-\sigma_B, \tau)} + (1-\epsilon)\delta_{(\sigma_B, \tau)}. \tag{B.11}$$

Onde, $\sigma_B = \text{sign}(\boldsymbol{\omega}^T\boldsymbol{x} + \eta)$, sendo que η corresponde ao ruído aditivo que corrompe a classificação com um ruído gaussiano ($\eta = \mathcal{N}(0, \sigma)$). Com isso, podemos reescrever (B.11) como,

$$p(\tau|\boldsymbol{\omega}, \boldsymbol{x}, \boldsymbol{\eta}) = \epsilon + (1 - 2\epsilon)\Theta\left(\tau\left(\boldsymbol{\omega}^T\boldsymbol{x} + \eta\right)\right). \tag{B.12}$$

Marginalizando (B.12) em relação ao ruído aditivo teremos que

$$p(\tau|\boldsymbol{\omega}, \boldsymbol{x}) = \epsilon + (1 - 2\epsilon)\frac{1}{\sqrt{2\pi\sigma^2}}\int \mathrm{d}\eta e^{-\frac{\eta^2}{2\sigma^2}}\Theta\left(\tau\left(\boldsymbol{\omega}^T\boldsymbol{x} + \eta\right)\right). \tag{B.13}$$

Com isso, podemos calcular a expressão para a evidência (B.10)

$$\begin{aligned}\mathcal{E} = \epsilon + (1 - 2\epsilon)&\frac{1}{|2\pi\boldsymbol{C}|^{\frac{1}{2}}\sqrt{2\pi\sigma}} \\ &\int \mathrm{d}\boldsymbol{u}\mathrm{d}\eta e^{-\frac{1}{2}\boldsymbol{u}^T\boldsymbol{C}^{-1}\boldsymbol{u} - \frac{\eta^2}{2\sigma^2}}\Theta\left(\tau\overline{\boldsymbol{\omega}}^T\boldsymbol{x} + \tau\boldsymbol{u}^T\boldsymbol{x} + \tau\eta\right).\end{aligned} \tag{B.14}$$

Definindo novas variáveis,

$$\tilde{\boldsymbol{u}} = \begin{pmatrix}\boldsymbol{u}\\ \eta\end{pmatrix}, \quad \tilde{\boldsymbol{\Sigma}} = \begin{pmatrix}\boldsymbol{C} & 0\\ 0 & \sigma^2\end{pmatrix} \quad \text{e} \quad \tilde{\boldsymbol{x}} = \begin{pmatrix}\boldsymbol{x}\\ 1\end{pmatrix}, \tag{B.15}$$

logo a expressão (B.14) poderá ser escrita como

$$\begin{aligned}\mathcal{E} &= \epsilon + (1 - 2\epsilon)\frac{1}{|2\pi\boldsymbol{\Sigma}|^{\frac{1}{2}}}\int \mathrm{d}\tilde{\boldsymbol{u}}e^{-\frac{1}{2}\tilde{\boldsymbol{u}}^T\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{u}}}\Theta\left(\tau b + \tau\tilde{\boldsymbol{u}}^T\tilde{\boldsymbol{x}}\right), \\ &= \epsilon + (1 - 2\epsilon)\Phi\left(\frac{\tau b}{\lambda}\right),\end{aligned} \tag{B.16}$$

onde $\lambda^2 = \sigma^2 + |\boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x}|$ e função cumulativa[10] $\Phi(x)$ da gaussiana $\mathcal{N}(0, 1)$,

$$\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\mathrm{d}te^{-t^2/2}. \tag{B.17}$$

Além disso, no caso em que a classificação do assunto dada pelo produto escalar entre o assunto e o parâmetro a ser estimado ($b_t = \overline{\boldsymbol{\omega}}_t^T\boldsymbol{x}_t$) , ou seja, $\tau = f(b)$, de acordo com a regra da cadeia, ($\frac{\partial f}{\partial\overline{\boldsymbol{\omega}}} = \boldsymbol{x}\frac{\partial f}{\partial b}$)

[10] Para fins computacionais, a função cumulativa $\Phi(x)$ da gaussiana $\mathcal{N}(0, 1)$ pode ser escrita em termos da função $\text{erf}(x) = (2/\sqrt{\pi})\int_{-\infty}^{x}\mathrm{d}t\exp(-t^2)$ através da relação $\Phi(x) = \frac{1}{2}(1 + \text{erf}(x/\sqrt{2}))$

as equações de aprendizado (B.9) podem ser rescritas como,

$$\begin{aligned}\overline{\boldsymbol{\omega}}_{t+1} &= \overline{\boldsymbol{\omega}}_t - \boldsymbol{C}_t \boldsymbol{x}_t \frac{\partial V_t}{\partial b_t},\\ \boldsymbol{C}_{t+1} &= \boldsymbol{C}_t - \boldsymbol{C}_t \boldsymbol{x}_t \boldsymbol{x}_t^T \boldsymbol{C}_t \frac{\partial^2 V_t}{\partial b_t^2}.\end{aligned} \tag{B.18}$$

Com isso, recuperamos os resultados obtidos para o aprendizado ótimo do perceptron booleano obtidos para ruídos multiplicativos[11] e aditivos[12].

[11] O. Kinouchi and N. Caticha. Learning algorithm that gives the Bayes generalization limit for perceptrons. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, 54(1):R54–R57, July 1996

[12] M. Biehl, P. Riegler, and M. Stechert. Learning from noisy data: an exactly solvable model. *Physical Review E*, 52(5):4624–4627, 1995

⌈B.4⌉ Passagens Matemáticas

Notas de Calculo Matricial

Seja $\boldsymbol{A}$ uma matriz inversível usaremos as seguintes propriedades de diferenciabilidade,

$$\begin{aligned}\mathrm{d}\boldsymbol{A}^{-1} &= -\boldsymbol{A}^{-1}\mathrm{d}\boldsymbol{A}\boldsymbol{A}^{-1},\\ \mathrm{d}\log\det\boldsymbol{A} &= \mathrm{tr}\left(\boldsymbol{A}^{-1}\mathrm{d}\boldsymbol{A}\right).\end{aligned} \tag{B.19}$$

Propriedade B.2.1

Seja $f_1(x)$ uma distribuição de probabilidade e $h(x)$ uma função bem comportada tal que,

$$\lim_{\boldsymbol{x}\to\pm\infty} h(\boldsymbol{x}) f_1(\boldsymbol{x}) = 0,$$

segue que,

$$\begin{aligned}0 &= \int \mathrm{d}\boldsymbol{x}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}}\left(h(\boldsymbol{x})f_1(\boldsymbol{x})\right),\\ 0 &= \int \mathrm{d}\boldsymbol{x}\mathrm{d}\left(h(\boldsymbol{x})f_1(\boldsymbol{x})\right),\\ &= \int \mathrm{d}\boldsymbol{x}\mathrm{d}h(\boldsymbol{x})f_1(\boldsymbol{x}) + \mathrm{d}f_1(\boldsymbol{x})h(\boldsymbol{x}),\\ &= E_1\left(\mathrm{d}h(\boldsymbol{x})\right) + E_1\left(h(\boldsymbol{x})\mathrm{d}\boldsymbol{x}^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)\right),\\ &= E_1\left(\frac{\mathrm{d}h(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}\right) + \boldsymbol{\Sigma}_1^{-1}E_1\left(\boldsymbol{x}h(\boldsymbol{x})\right) - \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 E_1\left(h(\boldsymbol{x})\right).\end{aligned} \tag{B.20}$$

No caso em que $f_1(\boldsymbol{x})$ é uma Gaussiana de média nula obtemos a propriedade,

$$E\left(\boldsymbol{x}h(\boldsymbol{x})\right) = \boldsymbol{\Sigma}_1 E\left(\frac{\mathrm{d}h(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}\right) = E(\boldsymbol{x}\boldsymbol{x}^T)E\left(\frac{\mathrm{d}h(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}\right). \tag{B.21}$$

Minimização KL para gaussianas - equação B.8

Seja $g(\boldsymbol{x})$ uma distribuição de probabilidade qualquer e $f_1(\boldsymbol{x})$ uma distribuição gaussiana dada por

$$f_1(x) = \frac{1}{|2\pi\boldsymbol{\Sigma}_1|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)} \tag{B.22}$$

logo a divergência KL entre essas duas distribuições será

$$D = \int \mathrm{d}\boldsymbol{x} g(\boldsymbol{x}) \log \frac{g(\boldsymbol{x})}{f_1(\boldsymbol{x})} \tag{B.23}$$

Agora, nosso objetivo é calcular os valores dos parâmetros de $f_1(\boldsymbol{x})$ que minimizam essa distancia . Primeiramente minimizamos em relação ao parâmetro $\boldsymbol{\mu}_1$ ($\frac{\mathrm{d}D}{\mathrm{d}\boldsymbol{\mu}_1} = 0$), Dessa forma temos,

$$\begin{aligned}
\mathrm{d}D &= \int \mathrm{d}\boldsymbol{x} g(\boldsymbol{x}) \mathrm{d}\log f_1(\boldsymbol{x}); \\
&= -\frac{1}{2}\int \mathrm{d}\boldsymbol{x} g(\boldsymbol{x}) \left(\mathrm{d}\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) + (\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}\mathrm{d}\boldsymbol{\mu}_1\right); \\
&= -\int \mathrm{d}\boldsymbol{x} g(\boldsymbol{x}) \mathrm{d}\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1).
\end{aligned} \tag{B.24}$$

Assim segue que,

$$\begin{aligned}
\boldsymbol{\mu}_1 &= \int \boldsymbol{x} g(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \\
&= E_g(\boldsymbol{x}).
\end{aligned} \tag{B.25}$$

Minimizando agora a divergência KL em relação à matriz de covariância $\mathbf{\Sigma}_1$ teremos

$$\begin{aligned}
\mathrm{d}D &= \int \mathrm{d}\boldsymbol{x} g(\boldsymbol{x}) \mathrm{d}\log f_1(\boldsymbol{x}), \\
&= -\frac{1}{2}\mathrm{d}\log|\mathbf{\Sigma}_1| - \frac{1}{2}\int \mathrm{d}\boldsymbol{x} g(\boldsymbol{x})(\boldsymbol{x}-\boldsymbol{\mu}_1)^T \mathrm{d}\mathbf{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1), \\
&= \mathrm{tr}\left(\Sigma_1^{-1}\mathrm{d}\mathbf{\Sigma}_1\right), \\
&\quad - \mathrm{tr}\left(\int \mathrm{d}\boldsymbol{x} g(\boldsymbol{x})(\boldsymbol{x}-\boldsymbol{\mu}_1)^T \mathbf{\Sigma}_1^{-1}\mathrm{d}\mathbf{\Sigma}_1\mathbf{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)\right), \\
&= \mathrm{tr}\left(\Sigma_1^{-1}\mathrm{d}\mathbf{\Sigma}_1\mathbf{\Sigma}_1^{-1}\left(\mathbf{\Sigma}_1 - \int \mathrm{d}\boldsymbol{x} g(\boldsymbol{x})(\boldsymbol{x}-\boldsymbol{\mu}_1)(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\right)\right), \\
&= \mathrm{tr}\left(\mathbf{\Sigma}_1^{-1}\mathrm{d}\mathbf{\Sigma}_1\mathbf{\Sigma}_1^{-1}\left(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_g\right)\right). \qquad \text{(B.26)}
\end{aligned}$$

Com isso, usamos o resultado $\mu_1 = E_g(x)$, obtemos que $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_g$.

Minimização KL - equações B.9 e B.10

Definindo $g(\boldsymbol{x}) = h(\boldsymbol{x}) f_0(\boldsymbol{x}-\boldsymbol{\mu}_0)/Z_h$ onde $Z_h = \int \mathrm{d}\boldsymbol{x} f_0(\boldsymbol{x}) h(\boldsymbol{x}) = E_0(h(\boldsymbol{x}))$ e $f_0(\boldsymbol{x})$ é uma normal de média $\boldsymbol{\mu}_0$ e covariância $\mathbf{\Sigma}_0$ Segue a partir de (B.27),

$$\begin{aligned}
\boldsymbol{\mu}_1 &= \frac{1}{Z_h}\int \boldsymbol{x} h(\boldsymbol{x}) f_0(\boldsymbol{x}-\boldsymbol{\mu}_0)\mathrm{d}\boldsymbol{x}, \\
&= \boldsymbol{\mu}_0 + \frac{1}{Z_h}\int \boldsymbol{u} h(\boldsymbol{u}+\boldsymbol{\mu}_0) f_0(\boldsymbol{u})\mathrm{d}\boldsymbol{u}, \\
&= \boldsymbol{\mu}_0 + \mathbf{\Sigma}_0\frac{1}{Z_h}\int \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{u}} h(\boldsymbol{u}+\boldsymbol{\mu}_0) f_0(\boldsymbol{u})\mathrm{d}\boldsymbol{u}, \\
&= \boldsymbol{\mu}_0 + \mathbf{\Sigma}_0\frac{1}{Z_h}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\mu}_0}\int h(\boldsymbol{u}+\boldsymbol{\mu}_0) f_0(\boldsymbol{u})\mathrm{d}\boldsymbol{u}, \\
&= \boldsymbol{\mu}_0 + \mathbf{\Sigma}_0\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\mu}_0}\log Z_h. \qquad \text{(B.27)}
\end{aligned}$$

Seja $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \boldsymbol{v}$ temos agora para a matriz de covariância que

$$\begin{aligned}
\boldsymbol{\Sigma}_1 &= \frac{1}{Z_h}\int \boldsymbol{x}\boldsymbol{x}^T h(\boldsymbol{x}) f_0(\boldsymbol{x}-\boldsymbol{\mu}_0)\mathrm{d}\boldsymbol{x} - \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T,\\
&= \frac{1}{Z_h}\int \boldsymbol{u}\boldsymbol{u}^T h(\boldsymbol{u}+\boldsymbol{\mu}_0) f_0(\boldsymbol{u})\mathrm{d}\boldsymbol{u} + \boldsymbol{\mu}_0\boldsymbol{\mu}_0^T + \boldsymbol{\mu}_0\boldsymbol{v}^T + \boldsymbol{v}\boldsymbol{\mu}_0^T - \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T,\\
&= \frac{1}{Z_h}\boldsymbol{\Sigma}_0\int \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{u}}\boldsymbol{u}^T h(\boldsymbol{u}+\boldsymbol{\mu}_0) f_0(\boldsymbol{u})\mathrm{d}\boldsymbol{u} - \boldsymbol{v}\boldsymbol{v}^T\\
&= \frac{1}{Z_h}\boldsymbol{\Sigma}_0\int \left(\mathbf{1}h(\boldsymbol{u}+\boldsymbol{\mu}_0) + \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{u}}h(\boldsymbol{u}+\boldsymbol{\mu}_0)\boldsymbol{u}^T\right) f_0(\boldsymbol{u})\mathrm{d}\boldsymbol{u} - \boldsymbol{v}\boldsymbol{v}^T,\\
&= \boldsymbol{\Sigma}_0 + \frac{1}{Z_h}\boldsymbol{\Sigma}_0\int \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{u}}h(\boldsymbol{u}+\boldsymbol{\mu}_0)\boldsymbol{u}^T f_0(\boldsymbol{u})\mathrm{d}\boldsymbol{u} - \boldsymbol{v}\boldsymbol{v}^T,\\
&= \boldsymbol{\Sigma}_0 + \frac{1}{Z_h}\boldsymbol{\Sigma}_0\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\mu}_0}\int h(\boldsymbol{u}+\boldsymbol{\mu}_0)\boldsymbol{u}^T f_0(\boldsymbol{u})\mathrm{d}\boldsymbol{u} - \boldsymbol{v}\boldsymbol{v}^T,\\
&= \boldsymbol{\Sigma}_0 + \frac{1}{Z_h}\boldsymbol{\Sigma}_0\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\mu}_0}\left(\int \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{u}^T}h(\boldsymbol{u}+\boldsymbol{\mu}_0) f_0(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\right)\boldsymbol{\Sigma}_0 - \boldsymbol{v}\boldsymbol{v}^T,\\
&= \boldsymbol{\Sigma}_0 + \frac{1}{Z_h}\boldsymbol{\Sigma}_0\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\mu}_0}\left(\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\mu}_0^T}Z_h\right)\boldsymbol{\Sigma}_0 - \boldsymbol{v}\boldsymbol{v}^T,\\
&= \boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_0\frac{\mathrm{d}^2\log Z_h}{\mathrm{d}\boldsymbol{\mu}_0\boldsymbol{\mu}_0^T}\boldsymbol{\Sigma}_0. \qquad \text{(B.28)}
\end{aligned}$$

Integral B.16

Dado a integral,

$$f(\boldsymbol{b}, c) = \frac{1}{(2\pi)^{\frac{p}{2}}}\int \mathrm{d}\boldsymbol{x}\exp(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x})\Theta\left(\boldsymbol{b}\boldsymbol{x}^T - c\right). \qquad \text{(B.29)}$$

Definindo a transformação de variáveis, $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x}$, onde,

$$\boldsymbol{B}^{-1} = \begin{pmatrix} \boldsymbol{b}/|\boldsymbol{b}| \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_p \end{pmatrix} \qquad \text{(B.30)}$$

e $\{\boldsymbol{b}/|\boldsymbol{b}|, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_p\}$ formam uma base autonormal.

Logo a integral (B.29) poderá ser escrita em termos da nova variável

como

$$
\begin{aligned}
f(\boldsymbol{b}, c) &= \frac{1}{(2\pi)^{\frac{p}{2}}} \int \mathrm{d}\boldsymbol{y} |\boldsymbol{B}^{-1}| \exp(-\frac{1}{2}(\boldsymbol{B}^{-1}\boldsymbol{y})^T(\boldsymbol{B}^{-1}\boldsymbol{y}))\Theta\left(\boldsymbol{b}(\boldsymbol{B}^{-1}\boldsymbol{y})^T - c\right), \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}} \int \mathrm{d}\boldsymbol{y} |\boldsymbol{B}^{-1}| \exp(-\frac{1}{2}\boldsymbol{y}^T(\boldsymbol{B}\boldsymbol{B}^T)^{-1}\boldsymbol{y})\Theta\left(\boldsymbol{b}\boldsymbol{b}^T y_1/|\boldsymbol{b}| - c\right).
\end{aligned}
\tag{B.31}
$$

Tendo em vista que $|B^{-1}| = 1$ e

$$
\boldsymbol{B}\boldsymbol{B}^T = \begin{pmatrix} 1 & \tilde{\mathbf{0}} \\ \tilde{\mathbf{0}} & \boldsymbol{I} \end{pmatrix}, \tag{B.32}
$$

segue,

$$
\begin{aligned}
f(\boldsymbol{b}, c) &= \frac{1}{(2\pi)^{\frac{p}{2}}} \int \mathrm{d}y_1 \mathrm{d}y_2' \ldots \mathrm{d}y_n' \exp\left(-\frac{1}{2}\sum_{i=2}^{p} y_i'\right) \exp\left(-\frac{y_1^2}{2}\right) \Theta\left(by_1 - c\right), \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}} (2\pi)^{\frac{p-1}{2}} \int_{-\infty}^{\infty} \mathrm{d}y_1 \exp\left(-\frac{y_1^2}{2}\right) \Theta\left(by_1 - c\right), \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}} (2\pi)^{\frac{p-1}{2}} \int_{-\infty}^{\infty} \mathrm{d}y_1 \exp\left(-\frac{y_1^2}{2}\right) \Theta\left(by_1 - c\right), \\
&= \frac{1}{(2\pi b^2)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \mathrm{d}y' \exp\left(-\frac{y'^2}{2b^2}\right) \Theta\left(y' - c\right), \\
&= \Phi\left(\frac{c}{b}\right).
\end{aligned}
\tag{B.33}
$$

[C] Dinâmica de Opinião

Nos últimos anos é crescente o interesse em modelagem física de sociedades. Em 2009, foi publicado um *Review of Mordern Physics* compilando alguns dos mais recentes avanços da pesquisa nessa área [21][1]. A compreensão dos mecanismos que levam à formação de **consenso** e **dissenso** em sociedades é uma das atuais áreas de interesse na pesquisa de dinâmica social. A ideia é determinar quando a dinâmica de uma sociedade de agentes interagentes, que podem escolher entre diferentes opções (por exemplo, opinião contra ou a favor determinado assunto, características culturais, etc) chegará ao consenso, i.e. a maior parte da sociedade partilhará a mesma opinião, ou a estados com dissenso. Também é de interesse, no caso em que existe localização espacial dos agentes, encontrar dinâmicas que levam à estados **polarizados**, ou seja, estados em que o consenso é apenas local.

[1] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646, May 2009

Na literatura da área, existem duas principais vertentes de modelos de agentes. A mais estudada é a de modelos de agentes com opiniões discretas. Na maioria, são modelos do tipo Ising onde só existem dois tipos de opinião. A outra vertente são modelos com opinião contínua, entre os quais podemos destacar os de modelos de agentes com o conceito de *confiança limitada* (*bounded confidence*) onde os agentes só interagem se eles tiverem a diferença de suas opiniões dentro de um certo limite [31, 62, 77].

Do primeiro grupo, dois exemplos de modelos conhecidos são o modelo do *Votante* [2], e modelo de *Sznajd* [3]. Neles, os agentes são representados por uma variável binária $s = \pm 1$ e apresentam uma dinâmica simples. No caso do modelo do votante, a cada passo da dinâmica um agente i escolhido aleatoriamente assume a opinião de um

[2] R. Holley and T. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, 3(4):643–663, 1975

[3] K. Sznajd-Weron and J. Sznajd. Opinion evolution in closed community. *International Journal of Modern Physics C*, 11(6):1157–1165, 2000

de seus vizinhos, também escolhido aleatoriamente. Já na dinâmica do modelo de Sznajd um par de vizinhos é escolhido ao acaso. Caso eles apresentem a mesma opinião, esse par irá impor essa opinião para seus vizinhos em comum. No entanto, se esse par tiver opiniões diferentes, cada agente irá impor sua opinião ao vizinhos do outro agente. Outro modelo de opiniões discreta é o modelo da *Regra da Maioria* [4]. Nesse modelo, os agentes também apresentam opiniões binárias. A dinâmica do modelo se inicia com uma fração p_+ de agentes de opinião $+1$ e uma $p_- = 1 - p_+$ de opinião -1. A cada passo da dinâmica, um grupo de r agentes é escolhido aleatoriamente como um grupo de discussão. O resultado da discussão será que todos os agentes do grupo assumirão a opinião majoritária. O tamanho do grupo r é sorteado aleatoriamente a cada passo da dinâmica. Além disso, se r for ímpar sempre existirá uma maioria, no entanto, se r for par no caso da metade dos agentes do grupo apresentarem uma opinião diferente da outra metade, o grupo todo será enviesado em $+1$ sem perda de generalidade.

Uma importante característica desses modelos é que suas dinâmicas levam a estados de consenso além de serem equivalentes a processos de Monte Carlo com temperatura nula. Eles foram estudados sobre diversos contextos, sendo aplicados a áreas que vão além de dinâmica social. Foram feitos estudos sobre suas convergências para diversos grafos, além de existirem generalizações com mais de duas opiniões [21]. Apesar de serem interessantes do ponto de vista de problemas de dinâmicas probabilísticas, esses modelos apresentam pouca fundamentação experimental, sendo construídos a partir de ideias muito gerais. Alguns dos problemas que esses modelos apresentam são a falta de memória de seus agentes, e no caso de modelos discretos, a impossibilidade de se medir o grau de similaridade entre dois indivíduos. Afim de solucionar esses problemas, foram sugeridos na literatura modelos de *agentes brownianos* [5] onde cada agente apresenta uma energia interna que é usada para influenciar o seu comportamento futuro. Outra abordagem é o modelo opinião contínua e ação discreta (*CODA*) [6], nesse modelo, o agente acredita em uma das duas opções

[4] S. Galam. Majority rule, hierarchical structures, and democratic totalitarianism: A statistical approach. *Journal of Mathematical Psychology*, 426434:426–434, 1986

[5] F. Schweitzer and J. Holyst. Modelling collective opinion formation by means of active Brownian particles. *The European Physical Journal B-Condensed ...*, 732:723–732, 2000

[6] A. Martins. Continuous opinions and discrete actions in opinion dynamics problems. *International Journal of Modern Physics C*, 19(4):617–624, 2008

possíveis com uma probabilidade que evolui no tempo de acordo com a regra de Bayes. O consenso da sociedade ocorre somente em nível local, aparecendo opiniões extremas e divergentes, caracterizando um estado polarizado.

Outra característica importante não considerada nos modelos de opinião binária é que pessoas discutem diversos assuntos simultaneamente, sendo muito deles não passíveis de representação por um único número. Para tanto, em 1997, Axelrod propôs um modelo de disseminação de cultura [7] bastante influente tanto na física quanto na sociologia. Seu modelo é uma generalização natural de modelos de dinâmica de opiniões com várias dimensões. Além disso, esse modelo incorpora os conceitos de *influência social* e *homofilia* que são duas características tidas como fundamentais entre os cientistas sociais para o entendimento da disseminação de cultura. A primeira, com explicamos na sessão 2.4, é a tendência de indivíduos se tornarem mais similares quando interagem e a segunda é a tendência de indivíduos parecidos interagirem mais frequentemente.

[7] R. Axelrod. The Dissemination of Culture A Model with Local Convergence and Global Polarization. *Journal of conflict resolution*, 41(2):203–226, 1997

Mais especificamente, os agentes são definidos por um conjunto de números inteiros e tamanho F , $(\sigma_1, \ldots, \sigma_F)$. Cada componente representa uma *característica cultural*(*features*), e cada característica pode assumir q valores, $\sigma_f = 0, 1, \ldots, q-1$. Os agentes estão situados sobre os nós de um grafo que pode ser regular ou não, e inicialmente suas características são sorteadas aleatoriamente. Na dinâmica do modelo, a cada passo de tempo, um agente i e um vizinho j são sorteados. A probabilidade desses dois agentes interagirem é dada pela média de características ω_{ij} comuns entre eles, que é definida por,

$$\omega_{ij} = \frac{1}{F} \sum_{f=1}^{F} \delta_{\sigma_f^i \sigma_f^j}$$

onde $\delta_{\sigma^i \sigma^j}$ é o delta de Kronecker. Na interação o agente j assume uma características do agente i que não é comum entre eles: $\sigma_f^j(t+1) = \sigma_f^i(t)$ se $\sigma_f^j(t) \neq \sigma_f^i(t)$. Os conceitos de homofilia e influência social são usados no modelo, pois quanto mais similares são os agentes maior a probabilidade deles interagirem e a interação entre os agentes os

deixam mais parecidos. A dinâmica do modelo de Axelrod é bastante rica, para um número pequeno de características é possível encontrar estados estacionários de consenso, no entanto, quando o número de características cresce o estado estacionário é de total dissenso.

⌈C.1⌉ Sociedade de Perceptrons

Nessa sessão é descrito o modelo proposto em [24] e que serviu de base para o trabalho principal apresentado na tese. Enfatizamos algumas propriedades desse modelo obtidas através de cálculos numéricos. Sobre o contexto de redes neurais, podemos dizer que os agentes do modelo são perceptrons com vetores sinápticos $\boldsymbol{\omega}_i \in \Re^N$, onde $i = 1, 2, \ldots, K$ é o índice do agente e K é número total de agentes na sociedade. A interação entre os agentes se dá pela discussão de assuntos públicos representados pelos vetores $\boldsymbol{x}_\mu$ sendo μ um índice temporal. A cada passo μ da dinâmica, uma dupla de agentes vizinho $\boldsymbol{\omega}_i$ e $\boldsymbol{\omega}_j$ é escolhida aleatoriamente para discutir um assunto $\boldsymbol{x}_\mu$ e com probabilidade $1/2$ um dos agentes é sorteado para ser o aluno enquanto o outro agente será o professor. Portanto, para cada interação um dos vizinhos tenta aprender a direção do outro usando o algoritmo da equação (C.1),

$$\begin{aligned} \boldsymbol{\omega}_i' &= \boldsymbol{\omega}_i^\mu - \frac{1}{N}\nabla_{\boldsymbol{\omega}_i} H + \boldsymbol{\eta}, \\ &= \boldsymbol{\omega}_i^\mu + \frac{1}{N}\sigma_j^\mu \boldsymbol{x}_\mu F_\delta + \boldsymbol{\eta}. \end{aligned} \tag{C.1}$$

Onde $H\left(\{\boldsymbol{\omega}_j\}_{j=1,\ldots,K}, \boldsymbol{x}\right)$ é a Hamiltoniana do sistema e $F_\delta = -\nabla_{\boldsymbol{\omega}_i} H$ é a função de modulação do aprendizado e $\sigma_j^\mu = sign\left(\boldsymbol{\omega}_j^\mu \cdot \boldsymbol{x}_\mu\right)$ é a classificação dada pelo agente j sobre o assunto que é discutido no tempo μ. O termo $\boldsymbol{\eta}$ é um ruído gaussiano com média nula e pretende descrever uma falha de comunicação entre os agentes. Vamos impor que o vetores sinápticos dos agentes sejam sempre normalizados para impedir que existam agentes com opiniões muito mais relevantes que de outros na sociedade:

$$\tilde{\boldsymbol{\omega}}_i = \frac{\boldsymbol{\omega}_i'}{|\boldsymbol{\omega}_i'|}.$$

Para um ruído gaussiano com variância suficientemente grande, podemos fazer uma dinâmica de Metropolis[8] para a evolução da sociedade. Impondo a condição de balanceamento detalhado para a interação entre os agentes, o agente i aceitará a mudança nos seus pesos sinápticos ($\boldsymbol{\omega}_i^{\mu+1} = \tilde{\boldsymbol{\omega}}_i$) se isso diminuir a energia de interação com a sua vizinhança, caso contrário ele mudará seus pesos sinápticos com uma probabilidade p dada por,

[8] M. Newman and G. Barkema. *Monte Carlo methods in statistical physics*. Oxford University Press, Oxford, 1999

$$p = e^{-\alpha\left(H_{(new)} - H_{(old)}\right)}, \tag{C.2}$$

onde α é denominado de pressão de pares e é usualmente interpretado como o inverso da temperatura do sistema em outros modelos de Mecânica Estatística [9].

[9] No texto principal chamamos o termo de pressão de pares de β ao invés de α.

A Hamiltoniana que rege a dinâmica da sociedade deverá ser construída como a somatória do potencial de interação entre os pares,

$$H = \sum_{(i,j)} V\left(\boldsymbol{\omega}_i^{\mu}, \boldsymbol{\omega}_j^{\mu}, \boldsymbol{x}_{\mu}\right), \tag{C.3}$$

onde (i, j) representa a soma sobre primeiros vizinhos. A vizinhança da sociedade é definida através de um grafo $\mathcal{G}$ que pode ser tanto uma rede regular quanto um grafo com distribuição de arestas aleatória.

Assim como foi feito no texto principal, estudamos a dinâmica de opinião da sociedade no caso em que os agentes discutem um único assunto, chamado *Zeitgeist*, em todos os passos da dinâmica, ou seja, $\boldsymbol{x}_{\mu} = \mathcal{Z}$ para todo μ, sendo que $|\mathcal{Z}| = 1$.

O potencial de interação entre dois agentes na sociedade que foi introduzido por Caticha e Renato[24] é,

$$V\left(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j, \mathcal{Z}\right) = -\frac{1+\delta}{2} h_i h_j + \frac{1-\delta}{2} |h_i||h_j|. \tag{C.4}$$

onde $h_i^{\mu} = \boldsymbol{\omega}_i^{\mu} \cdot \boldsymbol{x}_{\mu}$ é a *sobreposição* entre o vetor sináptico do agente i e *Zeitgeist* ou simplesmente a *opinião* do agente em relação ao *Zeitgeist*. Já a função de modulação que rege o aprendizado dos agentes no modelo

é definida por

$$W_i^\mu = W_\delta^\mu = 1 - (1-\delta)\,\Theta\left(-h_i^\mu \sigma_j^\mu\right). \tag{C.5}$$

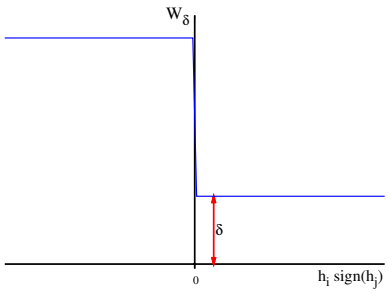
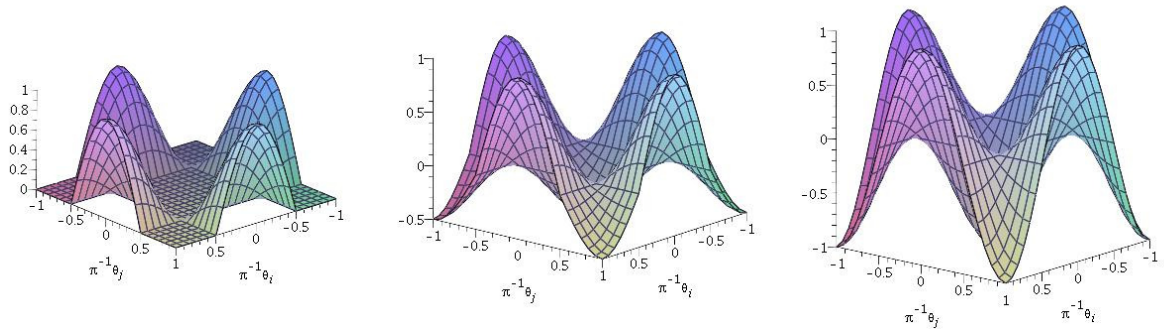


Figura C.1: Função de Modulação. O parâmetro δ representa a tendência que o agente tem de entrar em conformidade, ou seja, quanto maior for o valor de δ mais o agente irá aprender com um assunto que ele é concordante. Analogamente, podemos interpretar $1-\delta$ como a grau de importância à novidade.

Podemos perceber mais claramente através da figura C.1 que o parâmetro δ do modelo pode ser interpretado como uma tendência corroborativa dos agentes, pois mede a tendência do agente a mudar a direção do seu vetor sináptico quando ele está em concordância com seu parceiro social. A grandeza $1-\delta$ é interpretada como a importância que o agente dá a assuntos discordantes quando comparados com assuntos em que existe concordância com seus parceiros sociais. Essa interpretação também pode ser feita analisando a figura C.2 onde é apresentado o potencial de interação entre dois agentes para diferentes valores de δ.em função de θ_i e θ_j, com $h_{i,j} = \cos\theta_{i,j}$. É interessante notar que quando $\delta = 1$ essa função de modulação é idêntica a de Hebb e quando $\delta = 0$ ela é igual a função do perceptron de Rosemblatt[10].

[10] A. Engel and C. Van den Broeck. *Statistical mechanics of learning*. Cambridge University Press, Cambridge, 2004



(a) $\delta = 0.0$ (b) $\delta = 0.5$ (c) $\delta = 1.0$

Figura C.2: **Potencial de interação entre os agentes** para diferentes valores de δ em função de θ_i e θ_j, onde $h_{i,j} = \cos\theta_{i,j}$.

Para entendermos a dinâmica da sociedade acompanharemos o histogramas (q_h), médias (m) e variância (v) das opiniões dos agentes em relação ao *Zeitgeist* ($h_i = \boldsymbol{\omega}_i \cdot \mathcal{Z}$), que são definidos respectivamente por

$$q_h = \frac{1}{K}\sum_{i=1}^{K} I\left(h_i = h\right), \qquad h \in [-1,1], \tag{C.6}$$

$$m = \frac{1}{K}\sum_i h_i, \tag{C.7}$$

$$v = \frac{1}{K-1}\sum_i \left(m - h_i\right)^2, \tag{C.8}$$

$$\tag{C.9}$$

onde $I(x)$ é função indicação , que assume o valor unitário caso a asserção x seja verdadeira, e zero caso contrário.

Outra grandeza que pode ser usada para a análise do sistema é a **similaridade** entre as opiniões dos agentes, que definiremos como o produto escalar entre os vetores sinápticos entre dois agentes [11] $\rho_{ij} = \boldsymbol{\omega}_i.\boldsymbol{\omega}_j$. Acompanhamos o histograma (Q_ρ), média (M) e variância (V) dessa variável considerando todas as duplas de agentes, independentemente se eles são ou não vizinhos .

$$Q_\rho = \frac{1}{D}\sum_{i<j} I\left(\rho_{ij} = \rho\right), \qquad \rho \in [-1,1] \tag{C.10}$$

$$M = \frac{1}{D}\sum_{i<j}\rho_{ij}; \tag{C.11}$$

$$V = \frac{1}{D-1}\sum_{i<j}\left(M - \rho_{ij}\right)^2; \tag{C.12}$$

Onde $D = K\left(K-1\right)/2$ é o número total de duplas de agentes na sociedade. A vantagem de calcularmos essas grandezas é que, levando em consideração todas as duplas de agentes, obtemos uma flutuação menor quando comparadas com as medidas similares das opinião dos agentes. Isso possibilita uma melhor análise do perfil de opiniões dos agentes na sociedade sem que seja necessário tirar médias temporais ou sobre diferentes experimentos de monte carlo. Isso pode ser verificado atravéz da figura C.3, onde mostramos os histogramas da si-

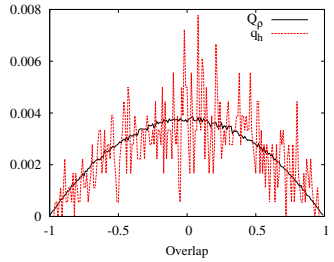


Figura C.3: Figura com as distribuições dos sobreposições entre agentes (Q_ρ) e do sobreposição entre os agentes e o assunto discutido q_h para $K = 900$ agentes com seus vetores sinápticos sorteados uniformemente sobre superfície de uma hiperesfera de $N = 5$ dimensões. Vemos que a distribuição Q_ρ apresenta menor flutuação que a distribuição (q_h) pois é calculada levando-se em conta a combinação sobre todas as duplas de agentes.

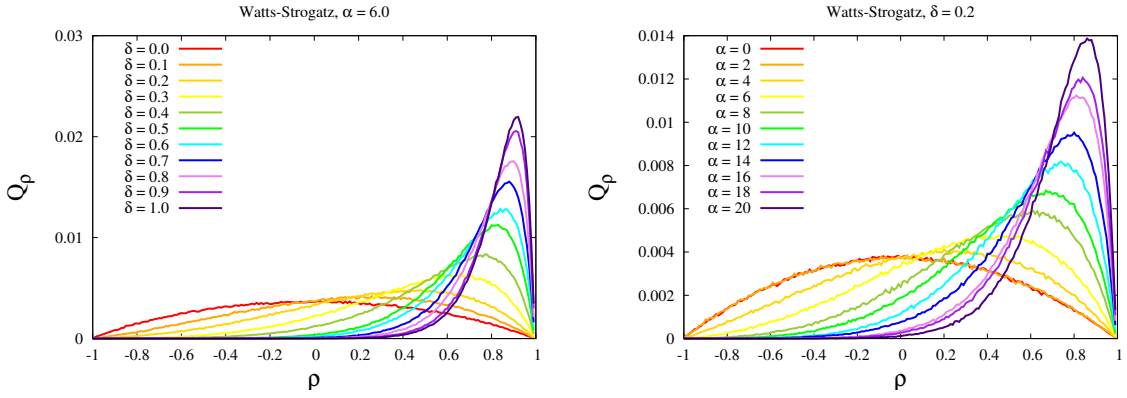
[11] Note que a variável ρ nesse contexto não tem nenhuma relação com o parâmetro de aprendizado a função de modulação do aprendizado do agente Bayesiano

milaridade e opiniões dos agentes quando os vetores sinápticos são distribuídos uniformemente sobre uma hiperesfera de 5 dimensões.

[C.2] Dependência com os parâmetros do modelo e diagramas de Fase

No modelo existem alguns parâmetros livres, os dois mais óbvios são a pressão de pares α e a tendência corroborativa do agente δ. Outros parâmetros livres são os parâmetros usados para definir o grafo $\mathcal{G}$ de suporte da sociedade. Entre eles, o número médio de vizinhos e o número total de agentes são os mais importantes.

Primeiramente vamos avaliar a influência dos parâmetros α e δ no estado de equilíbrio, olhando para os histogramas da similaridade dos agentes mostrados na figura C.4. Na figura C.4(a) fixamos o parâmetro de pressão $\alpha = 6.0$ e calculamos o histograma Q_ρ para vários valores da tendência corroborativa dos agentes, sendo que quanto mais vermelho menor o valor de δ. Com isso, podemos ver claramente, que quanto maior é a tendência corroborativa dos agentes mais os histogramas se alinham para valores com uma maior similaridade entre os agentes. Um efeito semelhante pode ser visto na figura C.4(b), onde fixamos o valor da tendência corroborativa dos agentes $\delta = 0.2$ e calculamos os histogramas para diversos valores da pressão entre pares α. Vemos que quanto maior o valor de α mais similares são os agentes.



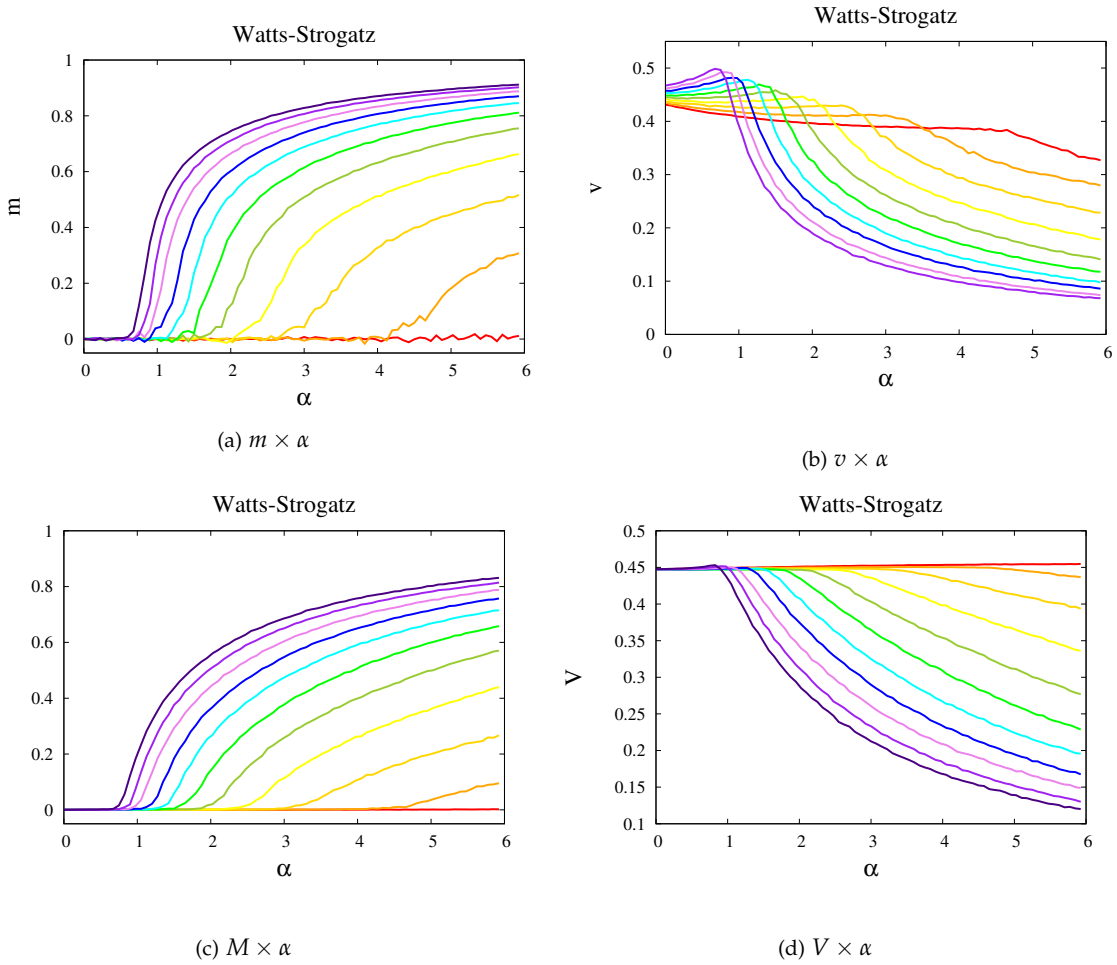
(a) $\alpha = 6.0$ (b) $\delta = 0.2$

Figura C.4: Histogramas das similaridades entres os agentes Q_ρ.

É importante notar na figura C.4 que para alguns valores pequenos dos parâmetros α e δ o histograma Q_ρ é o mesmo de uma sociedade totalmente desordenada que foi apresentado na figura C.3. Esse fenô-

meno pode ser observado com mais clareza nas figuras C.5(a) e C.5(c) [12] onde mostramos as médias das opiniões e das similaridades dos agentes em função da pressão de pares α, para vários valores de δ. Vemos claramente que existe uma região onde $m = 0$ e $M = 0$, sendo que para um valor de $\alpha_c(\delta)$ essas quantidades ficam maiores que zero. No caso da média (m) das opiniões dos agentes, essa mudança é abrupta o que caracteriza uma transição de fase de segunda ordem. Isso também pode ser verificado pois o gráfico C.5(d) onde apresentamos o gráficos da variância das opiniões dos agentes, que sofrem variações abruptas em suas derivadas na mesma região de parâmetro onde a opinião média se torna positiva. O fenômeno da transição de fase nesse sistema também foi estudado nas referências [24, 112].

[12] A escala de cor da figura C.5 é a mesma que da figura C.4(a)



(a) $m \times \alpha$

(b) $v \times \alpha$

(c) $M \times \alpha$

(d) $V \times \alpha$

Figura C.5: Gráficos com os valores das médias m , M e variâncias v, V de equilíbrio em função do parâmetro α para vários valores de δ. simulações, numa rede Watts-Strogatz e $K = 900$ agentes com número médio de vizinhos $n = 4.0$, média de uma amostragem 20 experimentos.

Nas figuras C.6 e C.7, apresentamos linhas de transição de fase que foram calculadas marcando os pontos nos quais a similaridade média dos agentes supera um pequeno valor de limiar ($M > \epsilon$).

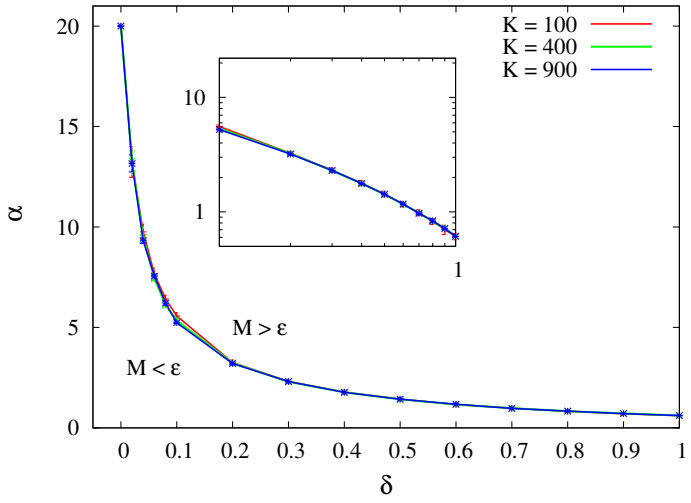
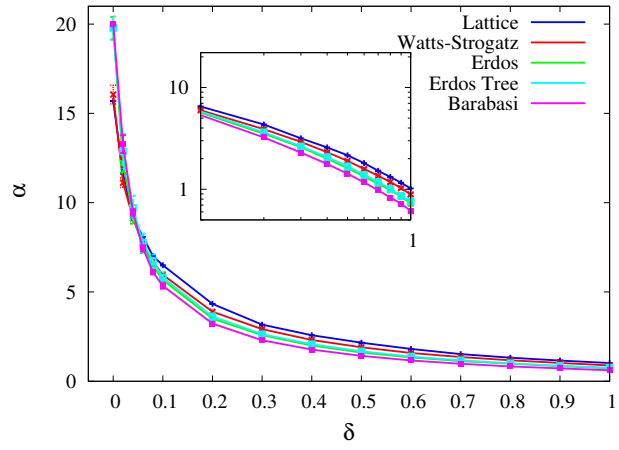


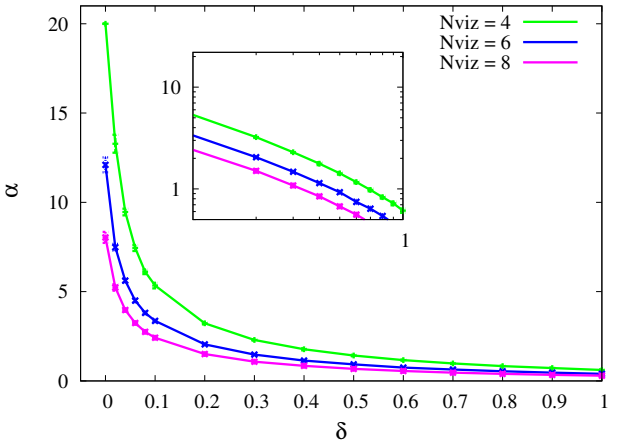
Figura C.6: Diagramas de fase para sociedades de tamanhos diferentes. A curva separa uma região de parâmetros que viabiliza o consenso da sociedade, sendo que a parte de baixo não viabiliza. Consideramos que existia consenso se o sobreposição médio entre as opiniões dos agentes fossem maior que um limiar ($M > \epsilon$). Calculamos esse gráfico com os agentes em uma rede do tipo Barabasi-Albert com $K = 900$ e com número médio de vizinhos $n = 4$. Para $K = 900$ agentes os pontos e suas barras de erro foram obtidos através de 10 simulações independentes, para $K = 400$ usamos 15 simulações e para $K = 100$ usamos 30 simulações.

Na figura C.6 calculamos o diagramas de fase para uma sociedade distribuída numa rede Barabasi-Albert com diferentes tamanhos. Podemos observar que a linha da transição de fase não depende do número total de agentes. Já na figura C.7(a) calculamos o diagrama de fase para sociedades com 900 agentes distribuídos sobre diferentes grafos de suporte. Também observa-se através desse gráfico que existe pouca dependência da região de transição de fase com a estrutura topológica da sociedade.

Por fim, avaliamos através da figura C.7(b) a dependência da transição de fase com o número médio de vizinhos numa rede do tipo Barabasi-Albert com 900 agentes. Como já é esperado a transição de fase ocorre para valores mais baixos do parâmetro de pressão social já que a energia de interação de uma agente com sua vizinhança cresce como o número de vizinhos.



(a)



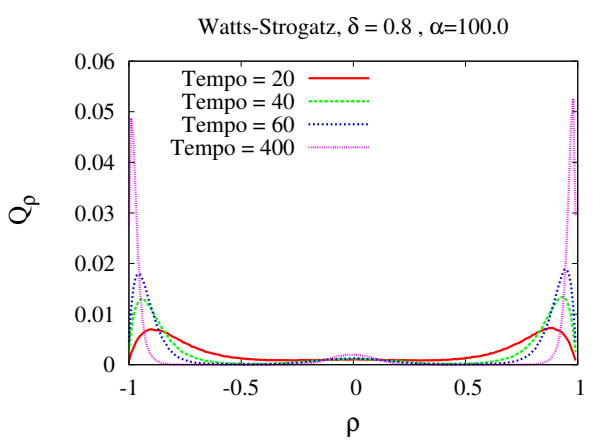
(b)

Figura C.7: Diagramas de fase calculados para diferentes redes e com diferentes números médios de vizinhos: As figuras acima foram calculadas com $K = 900$ agentes e os pontos e suas barras de erro foram feitos a partir de 20 simulações. Na figura (a) os diagramas de fase foram calculado para diferentes grafos. Percebemos através desse gráfico que a estrutura da sociedade tem pouca interferência sobre o comportamento dessa transição fase. Já na figura (b), calculamos o diagrama de fase, com $K = 900$ agentes dispostos sobre uma rede Barabasi-Albert, para diferentes números médio de vizinhos por sítio. Percebemos, por esse gráfico, que quanto maior o número médio de vizinhos, menor é o parâmetro α para o qual ocorre a transição de fase. Isso ocorre devido ao fato de que quanto mais vizinhos um agente tem, maior é a variação da energia causada por uma mudança no seu vetor de opinião.

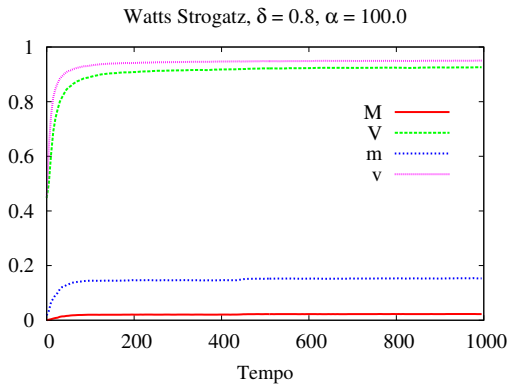
[C.3] Consenso e Polarização

O parâmetro δ da função de modulação, apesar de representar uma tendência natural do agente entrar em conformidade com seus vizinho, dá origem a uma importante característica desse modelo de opinião, que é a sua capacidade de criar estados polarizados. Essa característica foi pela primeira vez evidenciada na referência [118][13] para o caso em que o grafo de suporte da sociedade é um anel. Neste trabalho, a dinâmica do aprendizado dos agentes é feita através do algoritmo de aprendizado (C.1) mas sem incorporar nenhum ruído ao sistema. No nosso cenário de aprendizado, isso equivale a uma dinâmica de Monte Carlo com uma pressão social infinita ($\alpha \to \infty$). É importante salientar que os resultados apresentados nesta secção são obtidos partir de simulações onde a condição inicial dos vetores sinápticos de cada agente é sorteada uniformemente em uma hiperesfera de 5 dimensões.

[13] R. Vicente, A. C. R. Martins, and N. Caticha. Opinion dynamics of learning agents: does seeking consensus lead to disagreement? *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03015, Mar. 2009



(a) Distribuição Q_ρ calculada para vários tempos



(b) Evolução temporal das médias e variâncias dos sobreposições

Figura C.8: Figuras com a evolução temporal do processo de polarização da sociedade. Na figura (a) a sociedade se divide em duas facções uma distribuída em torno do *Zeitgeist* e outra na direção contraria. Já na figura (b) é apresentado a evolução temporal da média e variância do sobreposições. Os gráficos foram obtidos através de um experimento de Monte Carlo com $\alpha = 100$ e $\delta = 0.8$ para $K = 1089$ agentes que estão distribuídos sobre uma rede do tipo Watts-Strogatz, com número médio de vizinhos $n = 4$ e probabilidade de redistribuição de vértices $p = 0.1$.

Na figura C.8 mostramos uma rodada de Monte Carlo na qual a dinâmica do processo leva a um estado polarizado. Observa-se através dos histogramas Q_ρ para diferentes tempos de simulação que são apresentados na figura C.8(a), que a sociedade se divide em duas facções, uma que se distribui na direção do *Zeitgeist* e outra na direção contaria. Observa-se também, através da figura C.8(b), onde é apresentado

as médias e variâncias das opiniões e similaridades dos agentes, que a polarização da sociedade sobrevive para tempos longos de simulação. Essa simulação foi feita com alta pressão entre pares, $\alpha = 100$, e com agentes com grande tendência de entrar em conformidade, $\delta = 0.8$. Usamos $K = 1039$ agentes distribuídos sobre uma rede do tipo *Watts-Strogatz* (também conhecida como rede de mundo pequeno) com um número de vizinhos por sítio $n = 4$, e com um grau de rearranjo $p = 0.1$.

Podemos observar pela figura C.8(b) que, como a pressão social é muito grande, o sistema rapidamente atinge um estado de equilíbrio metaestável. Além disso, como $m \approx 0.15$ vemos que o tamanho dos domínios de agentes que estão na direção do *Zeitgeist* é um pouco maior que o domínio na direção contraria.

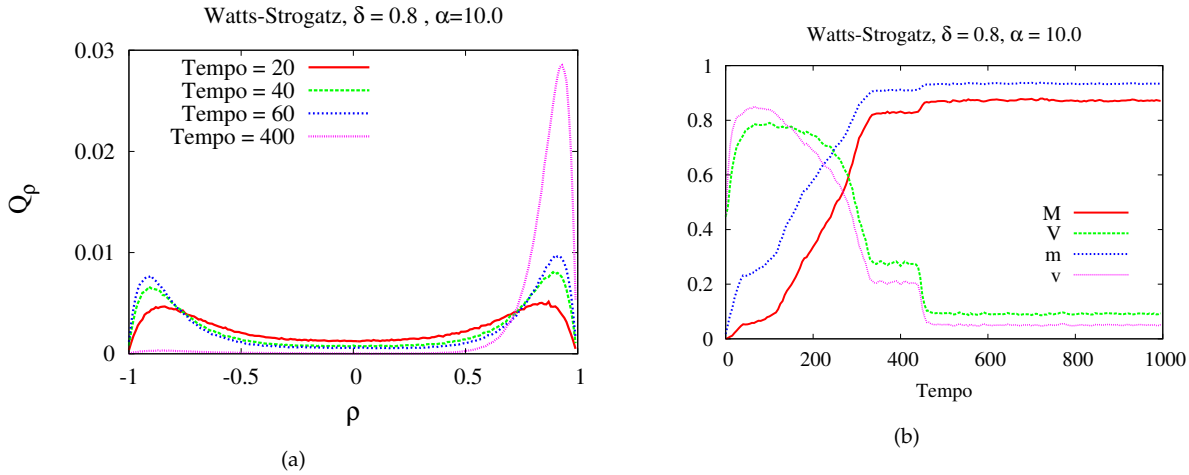
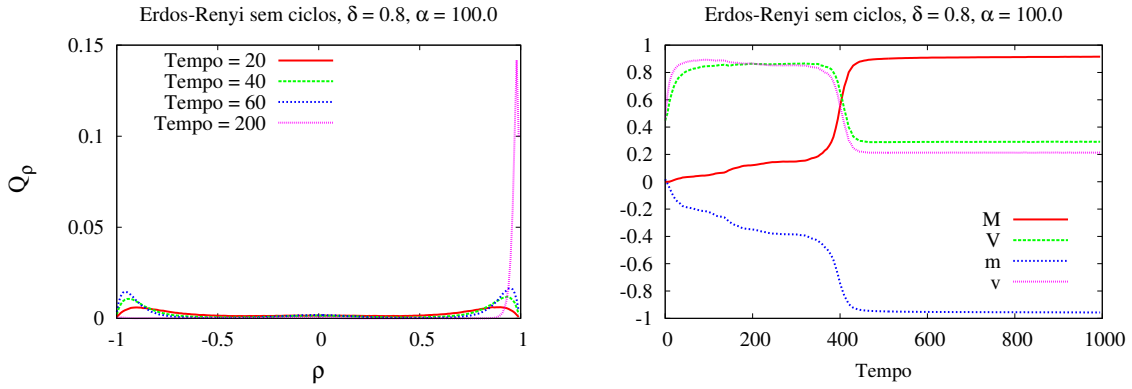


Figura C.9: Figuras com a evolução temporal da sociedade em uma dinâmica que leva a um estado consenso. São apresentados (a) as distribuições Q_ρ em diversos tempos. Inicialmente a dinâmica leva a estados polarizados. Com o passar do tempo uma das facções é destruída e o estado de equilíbrio é um estado de consenso. Na figura (b) é apresentado a evolução temporal das médias e variâncias das opiniões e similaridade dos agentes. A simulação é feita sob as mesmas condição das apresentadas na figura C.8, exceto o parâmetro de pressão social que é $\alpha = 10$.

No entanto, para pressões sociais menores as opiniões dos agentes são mais suscetíveis a flutuações. Isso faz com que a dinâmica da sociedade deixe de ser capaz de sustentar estados polarizados. Esse comportamento pode ser observado na figura C.9, que é obtida sobre as mesmas condições que a figura C.8, exceto pelo parâmetro do inverso da temperatura, que será $\alpha = 10$. Observa-se pela figura que inicialmente existe uma tendência de se criarem facções, no entanto, subitamente o sistema alcança o estado de equilíbrio que é o consenso.



(a) (b)

Figura C.10: Figuras com a evolução temporal da sociedade em uma dinâmica que leva a um estado consenso. Os gráficos acima foram obtidos a partir de uma uma rodada de simulação de Monte Carlo com $\alpha = 100$ e $\delta = 0.8$ para $K = 1089$ agentes que estão distribuídos sobre uma arvore aleatória.

Os mecanismos que levam esse sistema a estados polarizados precisam ainda ser melhor compreendidos. Sabemos que existe uma forte dependência deste fenômeno com a estrutura do grafo de suporte da sociedade e com condição inicial da distribuição de opiniões dos agentes sobre esse grafo. Na figura C.10 apresentamos o resultado de uma rodada de Monte Carlo feita sob as mesmas condições da figura C.8, com exceção do grafo de suporte da sociedade. Nessa simulação foi usado uma árvore aleatória ou grafo do tipo Erdös-Rényi sem ciclos. Vemos que mesmo com uma pressão social extremamente alta ($\alpha = 100$), a estrutura da sociedade não permite a manutenção de estados polarizados, pois em um curto intervalo de tempo a sociedade passa de um estado polarizado para um estado de consenso. É interessante observar que o estado de equilíbrio atingindo na dinâmica é um estado de consenso em que as opiniões dos agentes estão na direção contrária do assunto discutido ($m \approx -1$) . Esse tipo de comportamento é esperado devido ao fato da Hamiltoniana do sistema C.3 ser simétrica sobre a transformação $\omega_i \to -\omega_i$ para todo i.

[C.4] Diferentes Estratégias Cognitivas

Numa sociedade real, pessoas com diferentes estratégias cognitivas convivem entre si e suas interações alteram suas percepções sobre os mais diversos assuntos. Podemos tentar implementar esse fato em nossa modelagem da maneira mais direta possível; impondo que cada agente, com vetor de opinião $\boldsymbol{\omega}_i$, tenha uma estratégia cognitiva própria, representada por δ_i na sua função de aprendizado. Fazendo isso, o potencial de interação do agente i com um de seus vizinhos j será dado por.

$$V_{\delta_i}\left(h_i, h_j\right) = -\frac{1+\delta_i}{2} h_i h_j + \frac{1+\delta_i}{2}\left|h_i h_j\right| \tag{C.13}$$

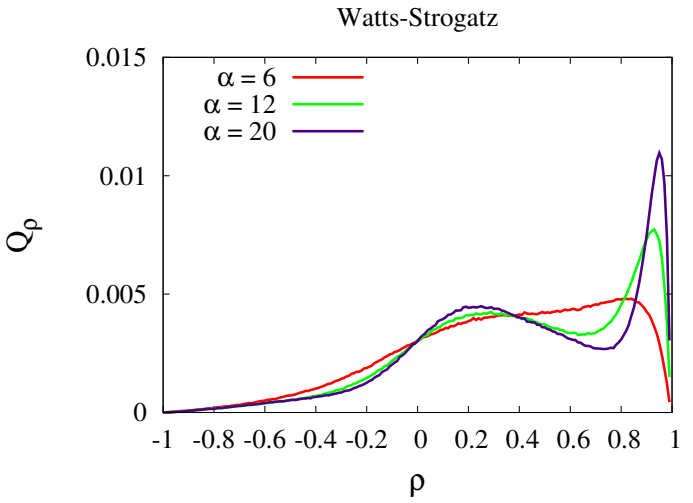
Na dinâmica, quando esse agente é escolhido ao acaso para interagir com seu vizinho, ele tentará aprender a opinião dele usando a equação C.1. Como foi feito anteriormente, o agente aceita a mudança na sua opinião se isso diminuir a energia interação com sua vizinhança, caso contrário ele aceita essa mudança com uma probabilidade

$$p_i = \exp\left(-\alpha \sum_{j \in \mathcal{V}_i}\left[V_{\delta_i}^{new}\left(h_i, h_j\right) - V_{\delta_i}^{old}\left(h_i, h_j\right)\right]\right), \tag{C.14}$$

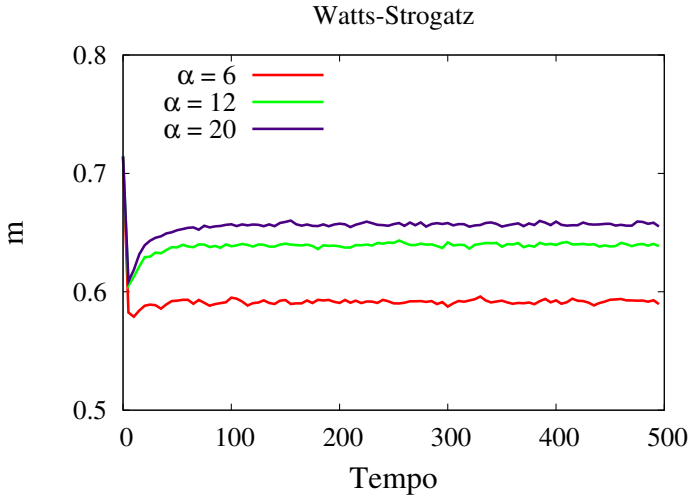
onde $\mathcal{V}_i$ é a vizinhança do agente i.

Por simplicidade, analisamos o caso extremo onde metade dos agentes têm a tendência a concordância máxima ($\delta = 1$) e a outra metade mínima ($\delta = 0$). A estratégia cognitiva do agente é sorteada independentemente da sua vizinhança com probabilidade $1/2$. Como a interação entre os agentes não é simétrica não podemos definir uma Hamiltoniana para o sistema, no entanto, isso não impede que a dinâmica do modelo leve a estados estacionários como observa-se nas figuras C.11 e C.12.

Na figura C.11(b) apresentamos a média das opiniões dos agentes sobre Zeitgeist (m) para diferentes valores do parâmetro α. Primeiramente, observa-se nessa figura que existe uma tendência de diminuição do consenso na sociedade, no entanto, esse comportamento é rapidamente revertido e a sociedade entra em equilíbrio num es-



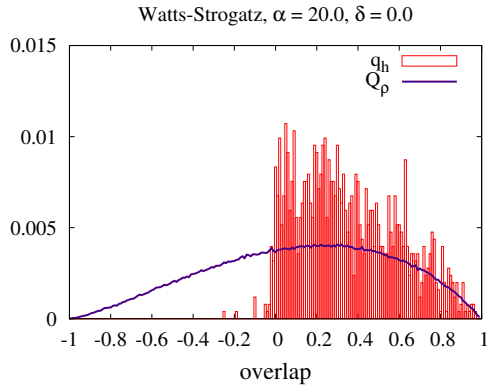
(a) Histograma sociedade completa



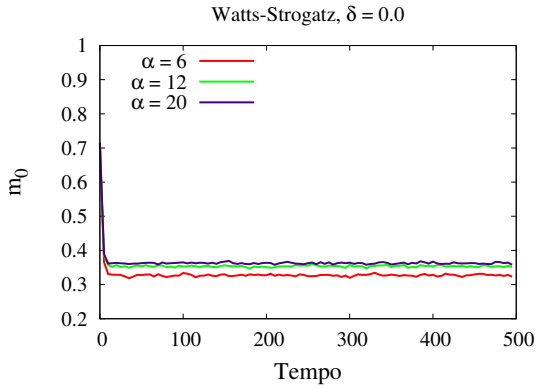
(b) Evolução temporal sociedade completa

Figura C.11: Figura com os histogramas e médias dos sobreposições entre agentes com a mesma estratégia cognitiva δ numa sociedade na qual metades dos agentes têm $\delta = 0$ e outra metade tem $\delta = 1$. Na figura (a) apresentamos os histogramas da similaridade entre todos os agentes (Q_ρ) para diferentes valores de pressão de pares α, enquanto em (b) apresentamos evolução temporal das médias das opiniões de todos os agentes (m). Essas figuras foram feitas com $K = 2500$ agentes com número médio de vizinhos $n = 4$ numa rede do tipo $Watts - Strogatz$. Os gráficos dos histogramas foram feitos usando uma rodada de simulação enquanto os gráficos da média das sobreposições foram construídos com 10 simulações.

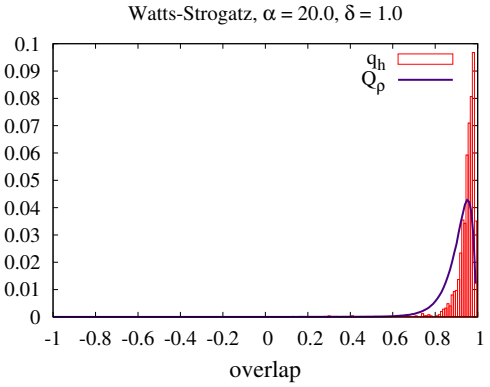
tado macroscópico de consenso que cresce em função do parâmetro α. Nesse cenário, vemos um comportamento curioso quando analisamos o histograma de Q_ρ para diferentes valores do parâmetro α, como é apresentado na figura C.11(a). Observa-se que à medida que α cresce surgem duas facções na sociedade, uma com bastante consenso e outra facção com bastante dispersão entre seu agentes.



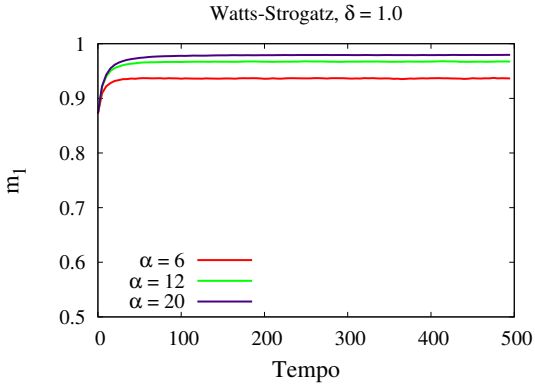
(a) Histogramas $\delta = 0$



(b) Evolução $\delta = 0$



(c) Histogramas $\delta = 1$



(d) Evolução $\delta = 1$

Figura C.12: sociedade com $K = 2500$ agentes sobre um grafo do tipo Watts-Strogatz com número médio de vizinhos $n = 4$. O histograma foi obtido usando uma rodada de simulação e o gráfico com a média dos sobreposições foi obtido através de 20 rodadas de simulação.

O efeito da interação entre agentes com estratégias cognitivas diferentes pode ser melhor avaliado a partir da figura C.12 onde apresentamos o histogramas Q_ρ e q_h e a média m do sobreposição para agentes que tem o mesmo tipo de estratégia cognitiva. Observa-se pelas figuras C.12(a) e C.12(b) que os agentes com $\delta = 0$, devido à interação com agentes de $\delta = 1$ se alinham entorno do *Zeitgeist*, pois para valores de pressão social menores que $\alpha = 16$ (ver figura C.7(a)) é esperado que o estado de equilíbrio de agentes com essa estratégia seja totalmente

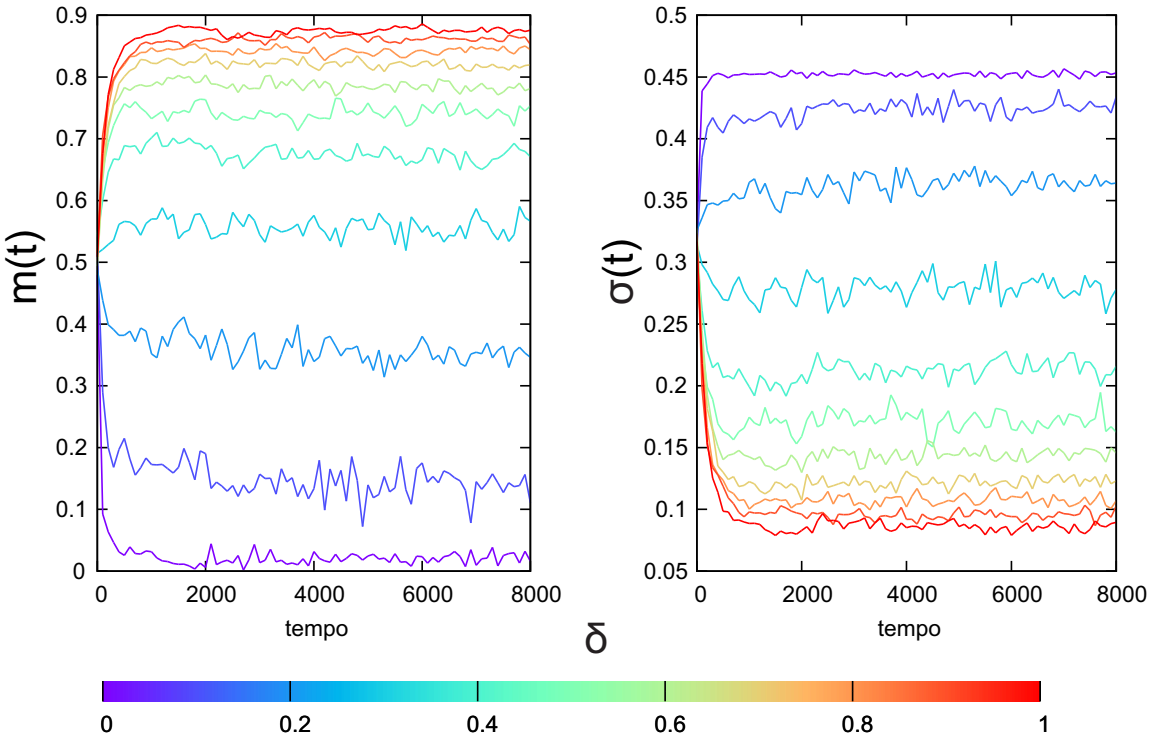
desorganizados, ou seja, $m = 0$. Vemos pelas figuras C.12(c) e C.12(d) que o efeito oposto o corre para agentes com $\delta = 1$ porém numa escala bem menor já esses agentes tem uma grande tendência a se alinharem ao redor do Zeitgeist. É importante salientar que a forma geral dos histogramas de opinião para agentes como a mesma estratégia cognitiva não sofre grandes modificações [14], no entanto, um estudo mais detalhado sobre o comportamento da sociedade de agentes com diferentes estratégias ainda precisa ser feito.

[14] Por exemplo, os histogramas não se tornam bimodais, ou sofrem mudanças mais drásticas

[C.5] Convencimento de Populações

No modelo de opinião que tratamos até aqui, os agentes sempre discutiam um assunto fixo, ao qual denominamos de *Zeitgeist*. Dessa forma, no estado de equilíbrio os agentes ficam distribuídos ao redor de uma direção fixa no espaço das opiniões. Vemos então que esse procedimento não nos permite avaliar como se dá a evolução do Zeitgeist na sociedade. Para tanto, iremos flexibilizar um pouco o modelo da seguinte maneira: após um tempo de termalização onde a sociedade discute um assunto fixo, o *Zeitgeist* discutido a cada passo da interação é calculado a partir das média dos vetores sinápticos dos agentes, ou seja,

$$\mathcal{Z}(t) \propto \frac{1}{K}\sum_i \omega_i(t) = \langle \omega_i(t) \rangle \, .$$



Com esse procedimento, esperamos que a direção do *Zeitgeist* entre em deriva. Para acompanharmos a dinâmica do modelo calculamos as médias e variâncias das opiniões dos agentes sobre o *Zeitgeist* instantâneo,

Figura C.13: Figura com a evolução temporal da média e variância entre o alinhamento do agentes e *Zeitgeist* para vários valores de δ e com a pressão social $\alpha = 8$, numa rede quadrada com 400 agentes. Até o tempo = 1000, a evolução temporal é feita com o *Zeitgeist* fixo, a partir desso ponto a evolução temporal é feita com o *Zeitgeist* calculado a partir das médias das opiniões. Percebemos que a evolução dessas grandezas permanece praticamente inalterada com a nova dinâmica.

$$m(t) = \langle \boldsymbol{\omega}_i(t) \cdot \mathcal{Z}(t) \rangle ,$$
$$\sigma(t) = \frac{1}{K} \left\langle (\boldsymbol{\omega}_i(t) \cdot \mathcal{Z}(t))^2 \right\rangle - \langle \boldsymbol{\omega}_i(t) \cdot \mathcal{Z}(t) \rangle^2 . \tag{C.15}$$

Essa modificação não muda de forma significativa a distribuição de opiniões do estado de equilíbrio em relação ao assunto discutido, como podemos ver na figura C.13 onde é mostrado a evolução temporal da média e variância entre o alinhamento do agentes e *Zeitgeist* instantâneo. Até o tempo = 1000, a evolução temporal é feita com o *Zeitgeist* fixo, a partir desso ponto a evolução temporal é feita com o *Zeitgeist* calculado a partir das médias das opiniões. Percebemos que a evolução dessas grandezas permanece praticamente inalterada com a nova dinâmica.

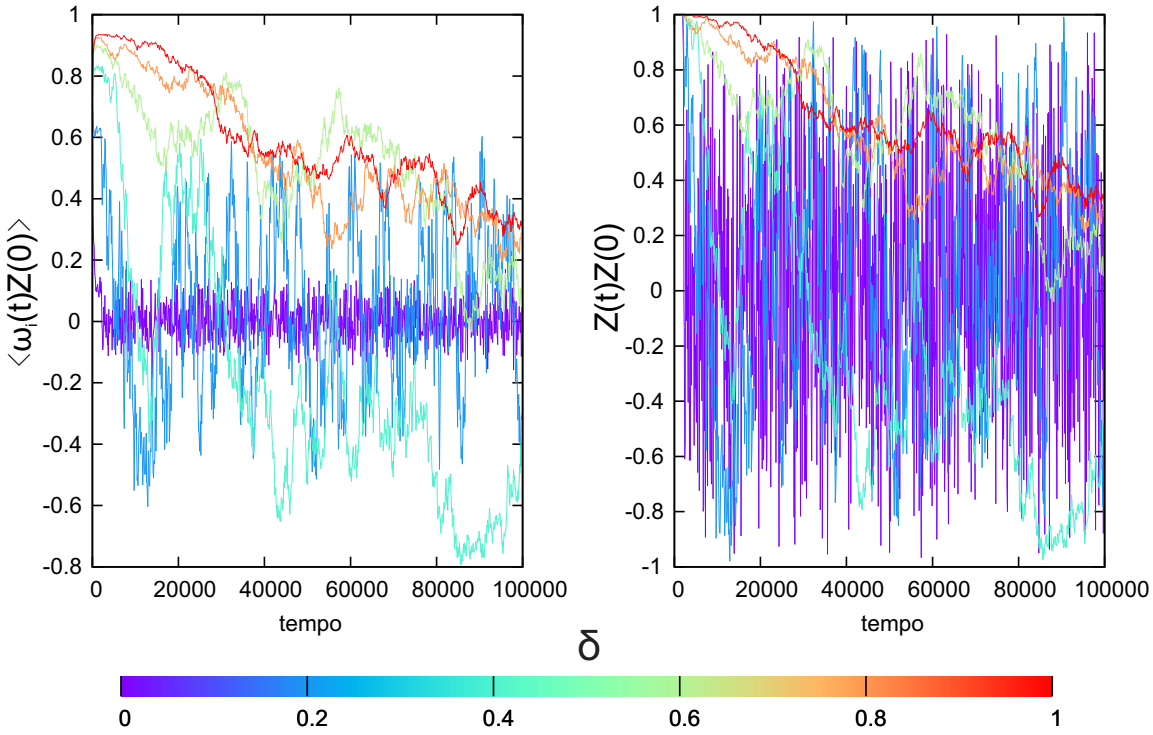


Figura C.14: Figura com movimento de deriva do *Zeitgeist*. É mostrado a média do alinhamento dos agentes com o *Zeitgeist* inicial, e o produto escalar entre o *Zeitgeist* inicial e o instantâneo. Observa-se nessa figura, que quanto menor o valor do parâmetro δ mais facilmente os agentes perdem a memória do *Zeitgeist* inicial. Figura simulação feita com 400 agentes dispostos sobre uma rede quadrada.

O movimento de deriva do *Zeitgeist* pode ser analisado através da figura C.14, onde mostramos na esquerda a média do alinhamento en-

tre os agentes e o *Zeitgeist* inicial $\langle \omega_i(t) \cdot \mathcal{Z}(0) \rangle$, e na direita o alinhamento entre o *Zeitgeist* instantâneo e o inicial $\mathcal{Z}(t) \cdot \mathcal{Z}(0)$. Podemos constatar que, de fato, o *Zeitgeist* apresenta um movimento de deriva e, além disso, vemos que quanto menor o valor do parâmetro δ mais facilmente os agentes perdem a memória do *Zeitgeist* inicial e mais acentuado são as flutuações das opiniões.

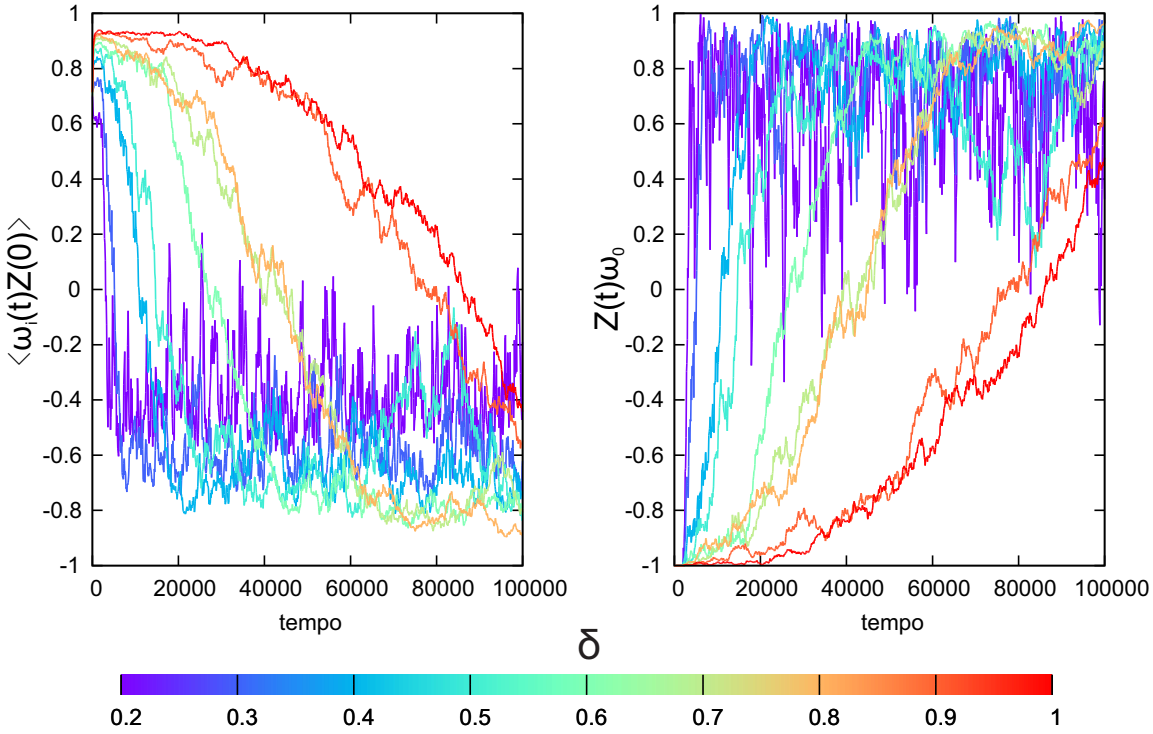


Figura C.15: Figura com a média dos alinhamentos entre as opiniões dos agentes e o *Zeitgeist* inicial, e o alinhamento entre o *Zeitgeist* instantâneo e a direção do agente fixo, que foi definida na direção oposta ao Zeitgeist inicial, $\omega_0 = -\mathcal{Z}$. Essa simulação foi feita sobre uma rede quadrada com 400 agentes interagentes e 1 agente fixo para vários valores do parâmetro δ

O primeiro passo que demos para estudar técnicas de convencimento de população foi supor que na população existia um único agente com vetor de opinião fixo e contrária à direção do *Zeitgeist* inicial $\omega_0 = -\mathcal{Z}(0)$. Esse agente não é parceiro social de nenhum outro agente, sua única influência na sociedade é no cálculo do *Zeitgeist*, ou seja,

$$\mathcal{Z}(t) \propto \sum_{i=1}^{k} \omega_i(t) + \omega_0. \tag{C.16}$$

Apesar da influência desse agente na sociedade ser muito pequena, sua presença na contabilização do *Zeitgeist* age como um pequeno campo externo que faz a sociedade convergir para sua direção em tempos longos. Podemos observar esse efeito na figura C.15 que mostra,

à esquerda, a média dos alinhamentos entre as opiniões dos agentes e o *Zeitgeist* inicial $\langle \boldsymbol{\omega}_i(t) \cdot \mathcal{Z}(0) \rangle$ e, à direita, o alinhamento entre o *Zeitgeist* instantâneo e a direção do agente fixo $\mathcal{Z}(t) \cdot \boldsymbol{\omega}_0$. Concluímos assim, que essa pequena modificação do modelo permite que estudemos futuramente a eficiência de métodos de convencimento populacional.

[C.6] Comparação entre ρ e δ

Para compararmos os parâmetros ρ e δ que definem as funções de modulação e características de aprendizado dos modelos estudados, iremos considerar por simplicidade o caso do aprendizado de um único agente no contexto de professor e aluno, e não o caso geral onde existem vários agentes interagentes.

Com isso, no limite termodinâmico, onde a dimensão dos assuntos discutidos entre o professor e alunos é muito grande, devido ao teorema do limite central [15] a distribuição da variável de concordância entre o professor e aluno $z = h\sigma$ segue uma distribuição normal de média zero e variância 1. Podemos definir uma função que mede o peso relativo da função de modulação de aprendizado entre assuntos concordantes e discordantes. Assim, para uma função de modulação genérica $G(z)$ a função de peso relativo pode ser definida por,

[15] A. Engel and C. Van den Broeck. *Statistical mechanics of learning*. Cambridge University Press, Cambridge, 2004

$$d_G = \frac{\int_0^\infty DzG(z) - \int_{-\infty}^0 DzG(z)}{\int_{-\infty}^\infty DzG(z)}. \tag{C.17}$$

No caso da função de modulação W_δ podemos mostrar que $d_\delta = \frac{1-\delta}{1+\delta}$. Assim, é simples isolar dessa expressão a variável δ em relação a função de peso relativo, de onde segue que,

$$\delta = \frac{1 - d_\delta}{1 + d_\delta}. \tag{C.18}$$

Generalizando essa expressão para uma função de modulação $G(z)$ obtemos uma expressão para a tendência corroborativa efetiva do agente,

$$\tilde{\delta} = \frac{1 - d_G}{1 + d_G}. \tag{C.19}$$

No caso da função de modulação do aprendizado Bayesiano 3.4,

calculamos numericamente a tendência corroborativa efetiva em função do parâmetro ρ. De acordo com o gráfico C.16 podemos observar que uma boa aproximação para relação entre esses dois parâmetros é $\tilde{\delta} \approx 1 - \rho$.

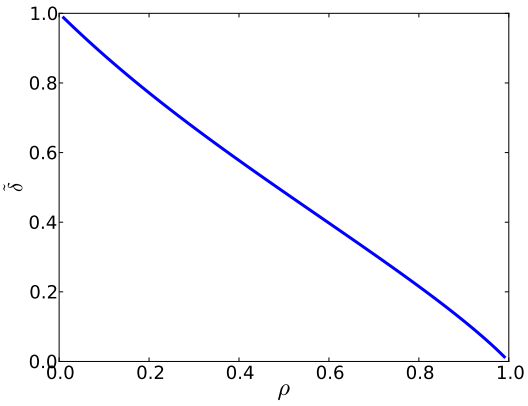


Figura C.16: Tendencia corroborativa efetiva $\tilde{\delta}$ em função do parâmetro de aprendizado ρ da função de aprendizado Bayesiano.

Bibliografia

[1] Asch conformity experiments - Wikipedia, the free encyclopedia.

[2] Morality Quiz/Test your Morals, Values & Ethics - Your Morals.Org.

[3] Neuron - Wikipedia, the free encyclopedia.

[4] D. Abrams, M. Wetherell, S. Cochrane, M. a. Hogg, and J. C. Turner. Knowing what to think by knowing who you are: self-categorization and the nature of norm formation, conformity and group polarization. *The British journal of social psychology / the British Psychological Society*, 29 (Pt 2)(May 1987):97–119, June 1990.

[5] R. Albert and A. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(January), 2002.

[6] J. Alford, C. Funk, and J. Hibbing. Are political orientations genetically transmitted. *American Political Science Review*, 99(2):153–167, 2005.

[7] D. M. Amodio, J. T. Jost, S. L. Master, and C. M. Yee. Neurocognitive correlates of liberalism and conservatism. *Nature neuroscience*, 10(10):1246–7, Oct. 2007.

[8] J. Arndt, J. Greenberg, S. Solomon, T. Pyszczynski, and L. Simon. Suppression, accessibility of death-related thoughts, and cultural worldview defense: exploring the psychodynamics of terror management. *Journal of personality and social psychology*, 73(1):5–18, July 1997.

[9] S. Asch. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, 70(9), 1956.

[10] R. Axelrod. The Dissemination of Culture A Model with Local Convergence and Global Polarization. *Journal of conflict resolution*, 41(2):203–226, 1997.

[11] A. G. Barto. Adaptive Critics and the Basal Ganglia Adaptive Critics and the Basal Ganglia. *Models of information processing in the basal ganglia*, page 215, 1995.

[12] J. W. Berry, Y. H. Poortinga, M. H. Segal, and P. R. Dasen. *Cross-cultural psychology: Research and applications*. Cambridge University Press, Cambridge, second edition, 2002.

[13] M. Biehl, P. Riegler, and M. Stechert. Learning from noisy data: an exactly solvable model. *Physical Review E*, 52(5):4624–4627, 1995.

[14] S.-J. Blakemore. The social brain in adolescence. *Nature reviews. Neuroscience*, 9(4):267–77, Apr. 2008.

[15] G. Bonanno and J. T. Jost. Conservative shift among high-exposure survivors of the September 11th terrorist attacks. *Basic and Applied Social Psychology*, 28(4):311–323, 2006.

[16] D. E. Brown. *Human universals.* Temple Publishe Press, Philadelphia, 1991.

[17] L. Buchen. Biology and ideology: The anatomy of politics. *Nature News*, 490:466, 2012.

[18] B. L. Burke, A. Martens, and E. H. Faucher. Two decades of terror management theory: a meta-analysis of mortality salience research. *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc*, 14(2):155–95, May 2010.

[19] D. K. Campbell-Meiklejohn, D. R. Bach, A. Roepstorff, R. J. Dolan, and C. D. Frith. How the opinion of others affects our valuation of objects. *Current biology : CB*, 20(13):1165–70, July 2010.

[20] W. Casebeer. Moral cognition and its neural constituents. *Nature Reviews Neuroscience*, 4(October):1–6, 2003.

[21] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646, May 2009.

[22] A. Caticha. Entropic Inference and the Foundations of Physics. *Brazilian Chapter of the International Society for Bayesian Analysis-ISBrA, Sao Paulo, Brazil*, 2012.

[23] N. Caticha, A. Susemihl, and R. Vicente. Diversity in cognitive styles leads to Culture Wars in an agent-based society. pages 1–26, 2010.

[24] N. Caticha and R. Vicente. Agent-Based Social Psychology: From Neurocognitive Processes To Social Data. *Advances in Complex Systems*, 14(05):711, 2011.

[25] J. F. Christensen and a. Gomila. Moral dilemmas in cognitive neuroscience of moral decision-making: a principled review. *Neuroscience and biobehavioral reviews*, 36(4):1249–64, Apr. 2012.

[26] R. B. Cialdini and N. J. Goldstein. Social influence: compliance and conformity. *Annual review of psychology*, 55(1974):591–621, Jan. 2004.

[27] R. T. Cox. Probability, Frequency and Reasonable Expectation. *American Journal of Physics*, 14(1):1, 1946.

[28] R. T. Cox. The Algebra of Probable Inference. *American Journal of Physics*, 31(1):66, 1963.

[29] C. T. Dawes and J. H. Fowler. Partisanship, Voting, and the Dopamine D2 Receptor Gene. *The Journal of Politics*, 71(03):1157, July 2009.

[30] F. B. M. de Waal. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Harvard University Press, 1997.

[31] G. Deffuant and D. Neau. Mixing beliefs among interacting agents. *Advances in Complex . . .* , (Ura 1306), 2000.

[32] M. H. DeGroot. *Probability and statistics*. Addison-Wesley, 1989.

[33] N. I. Eisenberger, M. D. Lieberman, and K. D. Williams. Does rejection hurt? An FMRI study of social exclusion. *Science (New York, N.Y.)*, 302(5643):290–2, Oct. 2003.

[34] A. Engel and C. Van den Broeck. *Statistical mechanics of learning*. Cambridge University Press, Cambridge, 2004.

[35] J. Epstein. *Generative social science: Studies in agent-based computational modeling*. Princeton University Press, Princeton, 2006.

[36] J. M. Epstein and R. Axtell. *Growing artificial societies : social science from the bottom up*. Brookings Institution Press., Washington DC, 1996.

[37] S. Fienberg. A brief history of statistics in three and one-half chapters: a review essay. *Statistical Science*, 7(2):208–225, 1992.

[38] A. P. Fiske. *Structures of social life: The four elementary forms of human relations: Communal sharing, authority ranking, equality matching, market pricing.*, volume 2. Free Press, New York, 1991.

[39] S. B. Flagel, J. J. Clark, T. E. Robinson, L. Mayo, A. Czuj, I. Willuhn, C. a. Akers, S. M. Clinton, P. E. M. Phillips, and H. Akil. A selective role for dopamine in stimulus-reward learning. *Nature*, 469(7328):53–7, Jan. 2011.

[40] T. Fø rland. Mentality as a social emergent: can the zeitgeist have explanatory power? *History and Theory*, 47(February):44–56, 2008.

[41] J. H. Fowler and D. Schreiber. Biology, politics, and the emerging science of human nature. *Science (New York, N.Y.)*, 322(5903):912–4, Nov. 2008.

[42] M. J. Frank, B. B. Doll, J. Oas-Terpstra, and F. Moreno. Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature neuroscience*, 12(8):1062–8, Aug. 2009.

[43] S. Galam. Majority rule, hierarchical structures, and democratic totalitarianism: A statistical approach. *Journal of Mathematical Psychology*, 426434:426–434, 1986.

[44] M. J. Gelfand, J. L. Raver, L. Nishii, L. M. Leslie, J. Lun, B. C. Lim, L. Duan, A. Almaliach, S. Ang, J. Arnadottir, Z. Aycan, K. Boehnke, P. Boski, R. Cabecinhas, D. Chan, J. Chhokar, A. D'Amato, M. Ferrer, I. C. Fischlmayr, R. Fischer, M. Fülöp, J. Georgas, E. S. Kashima, Y. Kashima, K. Kim, A. Lempereur, P. Marquez, R. Othman, B. Overlaet, P. Panagiotopoulou, K. Peltzer, L. R. Perez-Florizno, L. Ponomarenko, A. Realo, V. Schei, M. Schmitt, P. B. Smith, N. Soomro, E. Szabo, N. Taveesin, M. Toyama, E. Van de Vliert, N. Vohra, C. Ward, and S. Yamaguchi. Differences between tight and loose cultures: a 33-nation study. *Science (New York, N.Y.)*, 332(6033):1100–4, May 2011.

[45] A. S. Gerber, G. a. Huber, D. Doherty, C. M. Dowling, and S. E. Ha. Personality and Political Attitudes: Relationships across Issue Domains and Political Contexts. *American Political Science Review*, 104(01):111, Mar. 2010.

[46] A. E. Giannakakis and I. Fritsche. Social identities, group norms, and threat: on the malleability of ingroup bias. *Personality & social psychology bulletin*, 37(1):82–93, Jan. 2011.

[47] C. Gilligan. *In a Different Voice: Psychological Theory and Women"s Development*. Harvard University Press, Cambridge, MA, 1982.

[48] J. Graham, J. Haidt, and B. a. Nosek. Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, 96(5):1029–46, May 2009.

[49] J. Greenberg, L. Simon, T. Pyszczynski, S. Solomon, and D. Chatel. Terror management and tolerance: does mortality salience always intensify negative reactions to others who threaten one's worldview? *Journal of personality and social psychology*, 63(2):212–20, Aug. 1992.

[50] J. Greene. From neural 'is' to moral 'ought': what are the moral implications of neuroscientific moral psychology? *Nature reviews. Neuroscience*, 4(10):846–9, Oct. 2003.

[51] J. Greene and J. Haidt. How (and where) does moral judgment work? *Trends in cognitive sciences*, 6(12):517–523, Dec. 2002.

[52] J. D. Greene, L. E. Nystrom, A. D. Engell, J. M. Darley, and J. D. Cohen. The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44(2):389–400, Oct. 2004.

[53] J. D. Greene, R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen. An fMRI investigation of emotional engagement in moral judgment. *Science (New York, N.Y.)*, 293(5537):2105–8, Sept. 2001.

[54] J. Haidt. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108(4):814–834, 2001.

[55] J. Haidt. The new synthesis in moral psychology. *Science (New York, N.Y.)*, 316(5827):998–1002, May 2007.

[56] J. Haidt. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Pantheon Books, New York, 2012.

[57] J. Haidt and J. Graham. Planet of the Durkheimians, where community, authority, and sacredness are foundations of morality. In J. T. Jost, A. C. Kay, and H. Thorisdottir, editors, *Social and psychological bases of ideology*, volume Social and, chapter 15, pages 371–401. Oxford University Press, 2009.

[58] J. Haidt and C. Joseph. Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, (Special issue on human nature):55–66, 2004.

[59] J. Haidt and C. Joseph. The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In *The innate mind*, volume 3, chapter 19, pages 367–391. Oxford University Press, New York, 2007.

[60] J. Haidt and S. Kesebir. Morality. In *Handbook of Social Psychology*, chapter 22, pages 797–832. Wiley, 2010.

[61] P. K. Hatemi, N. a. Gillespie, L. J. Eaves, B. S. Maher, B. T. Webb, A. C. Heath, S. E. Medland, D. C. Smyth, H. N. Beeby, S. D. Gordon, G. W. Montgomery, G. Zhu, E. M. Byrne, and N. G. Martin. A Genome-Wide Analysis of Liberal and Conservative Political Attitudes. *The Journal of Politics*, 73(01):271–285, Jan. 2011.

[62] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Societies and Social Simulation*, 5(3), 2002.

[63] R. Holley and T. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, 3(4):643–663, 1975.

[64] C. B. Holroyd and M. G. Coles. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4):679–709, 2002.

[65] R. Iyer, S. Koleva, J. Graham, P. Ditto, and J. Haidt. Understanding libertarian morality: the psychological dispositions of self-identified libertarians. *PloS one*, 7(8):e42366, Jan. 2012.

[66] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, 2003.

[67] J. P. Jericó. *Aplicações de Mecânica Estatística a Sistemas Sociais : Interação e Evolução Cultural*. Mestrado, Universidade de São Paulo, 2012.

[68] J. T. Jost. The end of the end of ideology. *The American psychologist*, 61(7):651–70, Oct. 2006.

[69] J. T. Jost and D. M. Amodio. Political ideology as motivated social cognition: Behavioral and neuroscientific evidence. *Motivation and Emotion*, 36(1):55–64, Nov. 2011.

[70] J. T. Jost, J. Glaser, A. W. Kruglanski, and F. J. Sulloway. Political conservatism as motivated social cognition. *Psychological Bulletin*, 129(3):339–375, 2003.

[71] O. Kinouchi and N. Caticha. Optimal generalization in perceptrons. *Journal of Physics A: Mathematical and*, 25:6243–6250, 1992.

[72] O. Kinouchi and N. Caticha. Learning algorithm that gives the Bayes generalization limit for perceptrons. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, 54(1):R54–R57, July 1996.

[73] V. Klucharev, K. Hytönen, M. Rijpkema, A. Smidts, and G. Fernández. Reinforcement learning signal predicts social conformity. *Neuron*, 61(1):140–51, Jan. 2009.

[74] M. Koenigs, L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser, and A. Damasio. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, 446(7138):908–11, Apr. 2007.

[75] L. Kohlberg. Stage and sequence: The cognitive developmental approach to socialization. In *Handbook of socialization theory and research*. Rand McNally, Chicago, 1969.

[76] E. Kross, M. G. Berman, W. Mischel, E. E. Smith, and T. D. Wager. Social rejection shares somatosensory representations with physical pain. *Proceedings of the National Academy of Sciences of the United States of America*, 108(15):6270–6275, Apr. 2011.

[77] J. Lorenz. Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 18(12):1819–1838, 2007.

[78] N. G. Martin, L. J. Eaves, a. C. Heath, R. Jardine, L. M. Feingold, and H. J. Eysenck. Transmission of social attitudes. *Proceedings of the National Academy of Sciences of the United States of America*, 83(12):4364–8, June 1986.

[79] A. Martins. Continuous opinions and discrete actions in opinion dynamics problems. *International Journal of Modern Physics C*, 19(4):617–624, 2008.

[80] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087, 1953.

[81] J. Moll, R. Zahn, R. de Oliveira-Souza, F. Krueger, and J. Grafman. The neural basis of human moral cognition. *Nature Reviews Neuroscience*, 6:799–809, 2005.

[82] Y. Moriguchi, T. Ohnishi, T. Mori, H. Matsuda, and G. Komaki. Changes of brain activity in the neural substrates for theory of mind during childhood and adolescence. *Psychiatry and clinical neurosciences*, 61(4):355–63, Aug. 2007.

[83] P. R. Nail and I. McGregor. Conservative Shift among Liberals and Conservatives Following 9/11/01. *Social Justice Research*, 22(2-3):231–240, June 2009.

[84] P. R. Nail, I. McGregor, A. E. Drinkwater, G. M. Steele, and A. W. Thompson. Threat causes liberals to think like conservatives. *Journal of Experimental Social Psychology*, 45(4):901–907, July 2009.

[85] J. Neirotti and N. Caticha. Dynamics of the evolution of learning algorithms by selection. *Physical Review E*, 67(4):041912, Apr. 2003.

[86] M. Newman and G. Barkema. *Monte Carlo methods in statistical physics*. Oxford University Press, Oxford, 1999.

[87] H. Nishimori and G. Ortiz. *Elements of Phase Transitions and Critical Phenomena*. Oxford University Press, Oxford, 2011.

[88] M. Opper. On-line versus Off-line Learning from Random Examples: General Results. *Physical review letters*, 77(22):4671–4674, Nov. 1996.

[89] M. Opper and D. Saad. *Advanced mean field methods: Theory and practice*. The MIT Press, London, England, 2001.

[90] M. Opper and O. Winther. A Bayesian approach to on-line learning. In *On-line Learning in Neural Networks*. 1998.

[91] C. Panagopoulos. Social pressure, surveillance and community size: Evidence from field experiments on voter turnout. *Electoral Studies*, 30(2):353–357, June 2011.

[92] A. G. Patriota. A classical measure of evidence for general null hypotheses. *Fuzzy Sets and Systems*, pages 1–15, Mar. 2013.

[93] J. Piaget. *The Moral Judgment of the Child*. The Free Press, New York, 1965.

[94] D. a. Pizarro and P. Bloom. The intelligence of the moral intuitions: A comment on Haidt (2001). *Psychological Review*, 110(1):193–196, 2003.

[95] D. G. Rand, J. D. Greene, and M. a. Nowak. Spontaneous giving and calculated greed. *Nature*, 489(7416):427–430, Sept. 2012.

[96] P. Redgrave and K. Gurney. The short-latency dopamine signal: a role in discovering novel actions? *Nature reviews. Neuroscience*, 7(12):967–75, Dec. 2006.

[97] a. Rosenblatt, J. Greenberg, S. Solomon, T. Pyszczynski, and D. Lyon. Evidence for terror management theory: I. The effects of mortality salience on reactions to those who violate or uphold cultural values. *Journal of personality and social psychology*, 57(4):681–90, Oct. 1989.

[98] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386–408, Nov. 1958.

[99] S. H. Schwartz and W. Bilsky. Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology*, 58(5):878–891, 1990.

[100] F. Schweitzer and J. Holyst. Modelling collective opinion formation by means of active Brownian particles. *The European Physical Journal B-Condensed . . .*, 732:723–732, 2000.

[101] J. E. Settle, R. Bond, and J. Levitt. The Social Origins of Adult Political Behavior. *American Politics Research*, 39(2):239–263, Sept. 2010.

[102] J. E. Settle, C. T. Dawes, N. a. Christakis, and J. H. Fowler. Friendships Moderate an Association Between a Dopamine Gene Variant and Political Ideology. *The journal of politics*, 72(4):1189–1198, Jan. 2010.

[103] A. Shah. Psychological and Neuroscientific Connections with Reinforcement Learning. In M. Wiering and M. van Otterlo, editors, *Reinforcement Learning: State of Art*, chapter 16, pages 507–537. Springer-Verlag, Berlin Heidelberg, 2012.

[104] M. Sherif. An experimental approach to the study of attitudes. *Sociometry*, 1(1):90–98, 1937.

[105] N. J. Shook and R. H. Fazio. Political ideology, exploration of novel stimuli, and attitude formation. *Journal of Experimental Social Psychology*, 45(4):995–998, July 2009.

[106] R. A. Shweder, N. C. Much, M. Mahapatra, and L. Park. The "Big Three"of Morality (Autonomy, Community, Divinity) and the "Big Three"Explanations of Suffering. In *Morality and health*. 1997.

[107] K. B. Smith, D. R. Oxley, M. V. Hibbing, J. R. Alford, and J. R. Hibbing. Linking Genetics and Political Attitudes: Reconceptualizing Political Ideology. *Political Psychology*, 32(3):369–397, June 2011.

[108] S. Solla and O. Winther. Optimal perceptron learning: an online Bayesian approach. In *On-Line Learning in Neural Networks*. 1998.

[109] S. Solomon, J. Greenberg, and T. Pyszczynski. The Cultural Animal: Twenty Years of Terror Management Theory and Research. In *Handbook of experimental existential psychology*, chapter 2, pages 15–36. The Guilford Press, New York, 2004.

[110] L. H. Somerville, T. F. Heatherton, and W. M. Kelley. Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nature neuroscience*, 9(8):1007–8, Aug. 2006.

[111] J. Q. Stewart. The Development of Social Physics. *American Journal of Physics*, 18(5):239, 1950.

[112] A. K. Susemihl. *Aplicações de Mecânica Estatística à Psicologia Moral*. PhD thesis, Universidade de São Paulo, 2010.

[113] K. Sznajd-Weron and J. Sznajd. Opinion evolution in closed community. *International Journal of Modern Physics C*, 11(6):1157–1165, 2000.

[114] H. V. Tol, C. Wu, H. Guan, and K. Ohara. Multiple dopamine D4 receptor variants in the human population. *Multiple dopamine d4 receptor variants in the human population*, 358:149–142, 1992.

[115] M. Trusov, A. Bodapati, and R. Bucklin. Determining influential users in internet social networks. *Journal of marketing research*, XLVII(August):643–658, 2010.

[116] E. Tupes and R. Christal. Recurrent personality factors based on trait ratings. *Journal of personality*, 60(2):225–251, 1992.

[117] R. Vicente, O. Kinouchi, and N. Caticha. Statistical mechanics of online learning of drifting concepts: a variational approach. *Machine learning*, 201:179–201, 1998.

[118] R. Vicente, A. C. R. Martins, and N. Caticha. Opinion dynamics of learning agents: does seeking consensus lead to disagreement? *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03015, Mar. 2009.

[119] R. Vicente, A. Susemihl, J. a. P. Jericó, and N. Caticha. Moral foundations in an interacting neural networks society. *arXiv:1307.3203v1 - aguardando resposta da Physica A*.

[120] L. Wassermann. *All of statistics: A Concise Course in Statistical Inference*. Springer-Verlag, 2003.

[121] D. J. Watts. *Tudo É Óbvio - Desde Que Você Saiba a Resposta*. Paz e Terra, São Paulo, 2011.

[122] M. Weissflog, S. van Noordt, and B. Choma. Sociopolitical ideology and electrocortical responses. In *Psychophysiology*, volume 61, page 47, 2010.

[123] J. Woodward and J. Allman. Moral intuition: its neural substrates and normative significance. *Journal of physiology, Paris*, 101(4-6):179–202, 2008.

[124] N. Yeung, M. M. Botvinick, and J. D. Cohen. The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological Review*, 111(4):931–959, 2004.