学员专栏 · 黄倩盈

理解的结盟：从个体验证到关系构建

理解不是“我的模型对你猜得准不准”，而是“我们之间有没有立起一个双方都暂时受它约束的临时联盟”。理解的失败不是没猜准，是没入盟；理解的分析单元不在个体心里，在两个人之间那个受约束的关系场。

作者 黄倩盈 · SDE 学员 · 约 2.6 万字 · 发表于2026年7月22日

这篇文章的思想创新

“结盟”：理解是在二人关系场中建立一个相互约束的临时规范性联盟——双方暂时用同一套简化参照系、允许对方的信号约束自己的判断；理解的深度是联盟的可修正幅度，理解的失败是结盟行动的受挫，而非个体模型的误差。。

导读 · 300字

这篇文章的价值与意义 —— 它能解决你的什么问题

理论问题：它把理解的分析单元从个体内部搬到二人关系场，撤销“理解=跨越鸿沟去抵达对方内心”的验证隐喻。它与本栏《共同漫游》切的是相邻却不同的面：共同漫游问“关系本身如何成为一个以关系为本体的临时认知主体”，结盟问“这个关系里谁对谁负了什么可被违背、可被追责的规范性责任”——前者是本体涌现，后者是规范约束；共同漫游的产物随最后一个音符隐入无形，结盟的约束却在被背弃时产生确切的、可被经验到的后果。它进一步与共同基础、交互对齐、共享意图辨异：那些扭转了理解的认知／社会／互动条件，却没撤销那道“鸿沟”预设本身。

现实问题：它重述了当代“理解难”的病因——第一瓶颈已不是信道容量（技术给了空前的高带宽低延迟），而是结盟的发生条件正被三股力量系统性挤压：认知连续性被碎片化、低约束互动的充裕供应下调了默认预期、交互对象的加速可替换性摧毁了对初期挫败的容忍。结盟正从“默认会发生”变成“需要被有意识争取”的例外。

自我精神问题：它训练一种对“我进来了没有”的自省——当你追问对方“是工作上的事还是身体不舒服”，你可能仍站在自己房间里透过窗户猜；真正的入盟是那句放弃了用自己框架先一步组织对方经验的、完全开放的“今天发生了什么”。

摘 要

理解不是个体心智对另一个心智内部状态的逼近验证，而是一种旨在构建临时性认知联盟的动态过程。既有研究无论采取“管道隐喻”的信码传输，还是近年来的“互动校准”，都共有同一个未经审查的预设：理解最终发生在个体心智内部，是需要被个体获得的某种关于对方的“准确表征”。本文悬置这一预设，通过拆解一段日常对话中“理解失败”的深层结构，揭示出理解真正的要害不在“我的模型准不准”，而在“我们之间是否构建起了一个有约束力的临时联盟”——双方愿意暂时使用同一套简化的内部参照系，共享同一组未明言的注意优先级，并将后续互动建立在这一联盟之上。本文据此提出理解的分析单元应从个体内部转向“二人关系场”：理解的失败是结盟行动的受挫，理解的深度是联盟的可修正幅度。这一视角统一解释了面对面交谈中“自有预演”如何撕裂联盟、文本阅读中读者如何单方宣誓效忠、人机对话中联盟承诺的非对称性如何使理解的相互承认维度无法被激活。本文进一步论证，在一个交互资源充裕但注意力稀缺的时代，理解的第一瓶颈已不是信息传输的信道容量，而是结盟的发生条件正在被系统性地改变——那种愿意暂时悬置自己的预装网络、允许对方的信号对自己产生约束的动作，正在特定的注意力生态、风险结构和交互节奏中遭遇到结构性的发生困难。

关键词：理解；结盟；关系场；临时认知联盟；相互约束；发生学

一、引言：一个被搁置的追问

人与人交谈，读者与文本相遇，用户与人工智能对话——这三种经验我们都称之为“理解”。朋友听完倾诉后点点头，你觉得他懂了。读到三百年前的一行诗忽然怔住，你觉得那个死去的诗人站在你对面。对话框弹出几行字，精准抓住了你刚才描述的那个困境，你心想“它居然明白我在说什么”。三种场景，同一种“被理解”的感觉。于是很自然地，我们会追问：它们背后有没有一个共同的机制？

这一追问本身并没有错，但它早已被一条隐秘的岔路所捕获。这条岔路名为“验证隐喻”。它的基本逻辑是：理解即一个人内部的心理模型，对另一个人的内部心理状态，做出了足够准确的逼近。无论你把理解比喻为信息传输、共享知识库的建立、预测误差的最小化，还是临场临时理论的修正——这些理论彼此在细部上有重要差异，但它们共享同一个地基：理解被预设为一种“个体间”的验证事件，A要抵达B，所以要跨越鸿沟。理解的成败，就看桥搭得够不够牢。

本文要悬置的，正是这道鸿沟本身。它看起来像是一个关于理解的客观事实，但仔细一想，它其实是一个连问法都被规定了的问题设定。可万一理解从来不是A跨越鸿沟去抵达B，而是A与B共同决定，在两人之间暂时立起一个第三方的东西——一个临时的、双方都暂时承认的共享结构——那么，理解的成败就不在于谁的心智模型更准，而在于这个第三方东西有没有被建起来、双方有没有真的站在它下面。

这不是修辞。这是对理解发生条件的一笔根本改写。这一转向之所以被长期搁置，不是因为没有人触碰到它的边缘——事实上，常人方法学对日常互动中隐含预期的揭示（Garfinkel， 1967）、对话分析对修正序列的精细拆解、认知科学对交互对齐机制的建模（Pickering & Garrod, 2004）、行动哲学对共享意图与相互承诺的讨论（Bratman, 1992），都从不同方向逼近了同一个东西：理解不只是个体脑子里的事。特别是在 Herbert Clark（1996）的共同基础理论和 Michael Tomasello（2005）关于共享意图的发展心理学研究中，互动双方如何在对话中持续更新共享知识、如何从幼儿期就具备主动让渡认知主权的互动能力，这些机制已经被精细地描述过。但迄今为止，这些逼近始终未能正面撤销那道“鸿沟”预设本身。它们扭转了理解的认知条件、社会条件、互动条件，却没有追问那个最深的先验：理解的分析单元，到底应该落在个体内部，还是落在二人之间的关系场上？

本文的使命是将这一追问推向它的逻辑终点。本文将从一段真实的对话失败出发，解剖那道“鸿沟”预设无法解释的残余物——当修正程序已经启动、偏差已经暴露、信息通道已经打开时，理解为何仍然可能全盘搁浅？由这个裂缝入手，一个被既有理论触及却从未单独析出的维度将逐渐显形：理解中的相互承诺与相互约束关系。本文将这一维度的运作形态命名为“结盟”，并论证：结盟不是一个可以被还原为“共同基础”“临时理论”“交互对齐”或“共享意图”的新词，它是理解之为理解的那个不可替代的规范性内核。

二、对话失败的结构性解剖

以下片段改编自一段真实的伴侣争吵，细节已做调整以保护隐私，互动结构完整保留：

A（女性，下班回家，语气疲惫）：“我今天真的太累了。”

B（男性，正在看手机，头没抬）：“那你早点休息，别弄饭了，点外卖吧。”

A（停顿两秒，声音低了些）：“我不是说这个累……”

B（抬头，略带困惑）：“那你是什么意思？是工作上的事，还是身体不舒服？”

A（叹了口气）：“没什么，算了。”

最常见的分析会说：B误解了A的意思——A的“累”是一种需要被看见的情感耗竭，B却理解成身体疲劳。B的预装网络（累=身体疲劳=需要休息）成了一个糟糕的初始模型，后续也未能通过A的修正信号进行有效校准。这是“验证隐喻”的诊断：理解失败=模型误差没有被修正。

请仔细再看一遍这段对话的最后两句。A说“我不是说这个累”，这是她给B的一个明确的修正信号。B接收到这个信号，并且试图修正——他追问“是工作上的事，还是身体不舒服？”这个追问在形式上完全合乎“偏差暴露→修正”的规范程序。但A的反应是什么？她退出了。“没什么，算了。”

这是在验证隐喻里无法被充分解释的一个节点。如果理解只是模型修正，那B明明已经开始修正了，为什么A反而放弃了？你可以说，B的追问方式不对——它需要A进一步解释自己，而A太累了不想解释。但这个说法仍然停在验证隐喻的延长线上：它还是把失败归于B的认知操作质量——“修正程序虽然启动了，但算法不够聪明，没有一步到位”。

本文的诊断不同。A那句“我不是说这个累”暴露出的，不是B的模型不准，而是在那一瞬间，A意识到B并没有进入一个与她共享的参照系。这不是指B不知道A的内部状态，而是指A与B之间没有那件“我们正在同看的事”。A进来的时候，她的“累”不是一道待解的心理谜题，而是一个已经在现场的事实——就像两个人站在同一个房间里，应该都能看见同一把椅子。B的回应让她发现：他们不在同一个房间里。B正待在自己的房间里，试图透过一扇小窗猜测A的房间里有什么。

这里需要一个可操作的判准来区分两种不同的失败：“在房间里笨拙地摸索”与“根本不在房间里”。两者在行为表面可能高度相似——都表现为回应不准确，都触发了对方的修正信号——但在互动结构上有质的区别。入盟的摸索者会表现出对对方叙事优先级的主动让渡：他的追问不是给选项让A选，而是打开一个空白空间让A自己填充。他会问“今天发生了什么？”或“你刚才说的累，是指什么？”——这些问句的共同特征是不预设答案的类型，放弃用自己的框架先一步组织A的经验。未入盟者的追问则是框架强加：即使他摆出了选项，那些选项仍然是在他自己的预装网络内部生成的分类（身体不舒服/工作上的事），他没有在问句的构造本身里，给A的叙事留下一个不可被他的选项穷尽的开口。这个区分的核心不在于追问的内容是否“覆盖了正确的选项”，而在于追问的形式是否保留了对对方叙事主权的承认——即是否

在问句的语法内部就已经预判了答案的类型边界。

回到B的追问。“是工作上的事，还是身体不舒服？”这个问法从表面上看是在请求A帮助他校准模型。但恰恰是这个追问的方式，暴露了他仍然站在自己的参照系里：他仍然在用身体状态来框定A的疲惫（身体不舒服），然后试探性地加了一个选项（工作上的事），但对工作上的事到底是什么，他的问法本身没有给出任何信号表明他愿意进入那个具体的叙事。他没有表现出对“今天到底发生了什么”的探问兴致。A从他的追问里听出了这一点：他虽然问了，但他没有站进来。他不是在进入那件事，而是在自己的房间里，把窗户开大了一点。这正是框架强加的特征：B在A开口之前，已经将“疲惫”划入了“需要被归类到某一诱因类别中”的认知操作，这个操作本身就把A的疲惫当成了一道需要被解开的谜题，而不是一个需要被共同进入的叙事空间。

反过来，如果B真正入盟了，他可能会放下手机，停顿片刻，问：“今天发生了什么？”这句话的关键不在于它更“对”——对错仍然是认知维度的事。它的关键在于：它在宣告一种愿意进入A的参照系的姿态。它没有预设答案的类型（身体还是工作），它没有给A提供选项让其选择，它放弃了自己的框架先一步组织A经验的权利。这个放弃，就是入盟的核心动作：我暂时悬置自己的预装网络，让你那边的信号对我产生约束。这个问句在语法上是一个完全开放的邀请——它不包含任何预设的分类格，不暗示A必须把她的疲惫装进某个已有的抽屉。这种语法上的开放，就是对A叙事主权的承认的句法兑现。

这样一个入盟的姿态，并不保证会一步到位地理解A——B可能仍然会在后续的对话中犯错、需要纠正。但它保证了一件事：B已经站进了那个房间。他接下来犯的错误，将是发生在房间内部的错，而不是从房间外面透过窗户猜测的错。房间里发生的错是可以被A容忍的，因为A能感觉到B和自己正在同看一件事——他只是在看的过程中偏了，而不是根本没进来。房间内发生的错与房间外面发生的错，它们诱发的情感反应完全不同。A的退出，不是对“错”的反应，而是对“你不在房间里”的反应。这就是为什么验证隐喻永远无法解释A的“算了”：它只能分辨对与错，却分辨不了房内与房外。

A说“没什么，算了”，她放弃的不是“让B理解自己的累”——那仍然是把理解当成B要完成的认知任务。她放弃的，是与B共同建立那个房间。在验证隐喻的视野里，这只是一次修正失败的交互。但在另一种视野里，这是一个结盟行动的全盘搁浅。

三、理解即结盟：一个被搁置的基础动作

到目前为止，“结盟”这个词只是在隐喻的层面上被使用——它像是一个能照亮那段对话失败的意象，但还不是一个被严谨界定、能够承担论证重量的概念。本节的任务，不是在开篇给出一个精确定义然后去证明它，而是沿着前两节案例已经打开的经验裂缝，让这个概念从那些裂缝中逐步凝结出来。定义不是起点，而是追问的产物。

3.1 “我理解你”的实质动作

当一个人说“我理解你”，他通常不是在报告“我已对你建立了高精度的内部模型”。哪怕他对对方的内部状态所知甚少——比如对方刚刚经历了一场他并不清楚的失败——他仍然可以说“我理解”。这句话的实质动作，更像是一种宣告进入某种关系的言语行为：我愿意把你的事当成我的事，我愿意让你对我的判断产生约束，我把自己的一部分认知主权分给你。

这里的“分给”不是比喻。当B对A说“我理解你的累”，如果这不是敷衍，这句话就在履行一个确切的义务：从今往后，当他在另一个场合谈论A的累，或者在独处时回忆那段对话，他不能随意篡改A在那次谈话里给出的信号。A说的话，对他产生了一种他在独白时不会有的约束力。这个约束力，就是“结盟”的实质性内容。它不是在认知上建立高精度模型的结果，而是在关系上做出承诺的产物。

但“承诺”这个词本身也需要被进一步限定。在日常用法中，承诺往往意味着一种明确的、可以说出条款的约定——我向你保证做某事。结盟中的承诺，往往比这更轻、更不言明，但它仍然具备承诺的规范性内核：它改变了双方后续互动的许可范围。双方在日常的对话里，也是默认感知到这个范围的变化——哪些话可以在盟内说，哪些回应方式算作“背弃”，在大部分时候都不需要明说，但当背弃发生时，受损害的一方几乎立刻就能识别它。这就是规范性约束的核心特征：它不需要被事先书面化，但它在被违背时会产生确切的、可被经验到的后果。

约束力从何而来？这是一个必须被正面回答的元理论追问。凭什么“说一句话”就构成了承诺？在言语行为理论的传统中，承诺的效力通常依赖于制度背景、共同体承认或既有的关系结构。但结盟中的承诺极其轻便——它发生在两个陌生人之间（比如修车师傅和车主），几乎不依赖任何先在的制度框架，那么其规范性从何而来？

答案不在制度里，而在互动本身的内在规范性中。只要两个人进入了对话——不是擦肩而过的瞬间协调，而是开启了话轮交替、相互注视、彼此回应的对话序列——他们就已经默示地进入了一种相互责任的关系：我说的话是对你说的，我知道你在听，你知道我在对你说话。这个“对你”结构本身，就已经在互动的参与者之间织起了一层最薄的规范性织物。它不是从外部强加的规则，而是对话这种互动形式自身所携带的默示承诺：一旦你回应了，你就已经承认了对方的话对你产生了某种要求——至少，你不能在回应的同时完全不理会对方刚刚说了什

么。你可以违反这个要求，但违反的代价就是对话本身的瓦解。Garfinkel（1963）在讨论“信任”概念时已经触及了这个层面：日常互动依赖于一种“对场景可理解性的道德承诺”——参与者默认对方会按照互动的构成性规则行事，而这种默认本身具有规范性的约束力。结盟的规范性不是凭空产生的，它是从对话这种最基本的互动形式的道德承诺中延伸出来的：当你进入一段以理解为目的的对话，你不仅承诺了“我会按照对话规则行事”，你还进一步承诺了“我会暂时让你那边的信号对我产生约束”——这后一层是结盟特有的规范性增量，但它依赖的前一层（对话本身的默示承诺）是所有结盟可能发生的土壤。

3.2 结盟的所指物

这引出第一个关键的推论：结盟的所指物，从来不是“对方的内心状态”，而是“我们之间正在谈论的那件事”。A与B对话所结的盟，结在那个“疲惫”上，而不是结在B对A的认知模型上。当A说“我今天真的太累了”，她不是在邀请B猜她的心情，她是在邀请B进入一个二人共有的、关于“今天到底发生了什么”的叙事草稿。B的“点外卖”之所以失败，不是因为他的模型不准，而是因为他用自己的“身体疲劳脚本”另起了一件事，没有进入A要结的那个盟。A的后退，不是对修正程序的失望，而是对一个未被告知就解散的联盟的哀悼。

3.3 发生学追问：理解为什么必须采取“结盟”这种形态?

为什么理解必须采取“结盟”这种形态？是什么条件使单纯的个体认知操作不足以构成一次完整的理解发生?

一部分答案已经在第二节的分析中给出了：当两个人各自在自己的认知系统内运行，即使他们各自的模型运转得极其高效，他们之间仍然可能没有建立起那件“共同的事”。而“共同的事”之所以不可或缺，是因为理解所面对的初始材料——对方的话语、表情、姿态、沉默——在单独一个人的认知系统内，永远是开放性的。一个词、一个停顿、一个叹息，可以被填充进无数种不同的解读路径，而这些路径之间没有客观的、独立于双方互动之外的仲裁者。

在个体认知框架内，这种开放性通常被处理为“不确定性”——我需要更多的信息来缩小可能性空间。但这是一个错误的建模。它假定对方的内心状态已经是一个确定的、只待被更精确地逼近的东西，就好像博物馆里的展品已经摆好，只是灯光不够亮。实际上，在对话的初始时刻，连对方自己都未必对他的“内部状态”有一个清晰、固定的表征。人的情感和思想在进入语言之前，并不像展品那样边界分明——它们常常是模糊的、流动的、在一个表达行动中才第一次获得确定形态的。这就意味着，被理解的对象，不是在对话开始之前就完成、然后等待传输的东西；它在对话的过程中才逐步成形。

而这个“成形”的过程，本质上是双方共同参与的工作。A的疲惫，在她说出口之前，可能只是一团说不清的身体感受和情绪残渣。只有当B给出回应——一个具体的回应，带着具体的注意方向——这团东西才在二人的互动空间里逐渐显现出轮廓。B的每一次回应，不仅是在“猜测”A的状态，更是在参与构造那种状态的稳定形态。这就是为什么B站没站进房间这件事如此要紧：如果B站进去了，他那哪怕不完美的回应也是在这个共同空间中试探性地推进构造；如果B没站进去，那他的回应就只是在另一个空间里造一件不相干的东西。前者的构造可以被A承认、修改、共同完成；后者甚至无法被A认作是对自己那件事的回应。

从这个角度看，理解采取“结盟”这一形态，不是人类社会性的一种附加装饰，而是被理解的构造条件所逼出来的。理解的原始材料必须在一个共享空间里被共同赋予稳定形态，个体心智内部的模型处理只是这个共同构造过程的副产品——它们是在结盟成功之后沉淀在各自身上的残留物，而不是理解发生的原因。一个独自在脑中跑模型的人，无论跑得多准，都没有真正进入那件需要被共同构造的事。这就好比一个建筑师独自画完了一整栋楼的图纸，而那个本该与他一起盖楼的人从未出现在工地上——这栋楼从来没有被盖起来过。图纸的精度无关此事。

至此，“结盟”可以被给出一个正式的定义。这个定义不再是先验的起点，而是前文的经验分析和构造条件追问所逼出的必然结论：一次理解的发生，其最小条件不是单个个体获得了准确的内部表征，而是两个（或多个）参与者之间，建立起了一个临时性的认知联盟——双方愿意在同一段时间内，使用同一套简化的内部参照系（包括共享的术语、默认的注意方向、被共同承认的可推论的义务与边界），来展开后续的互动性构造，并承认这一构造过程对双方后续行动具有约束力。 这一定义将理解的分析单位从“个体内部”移到了“二人关系场”。理解不是A得到的东西，而是A与B之间发生的东西。

3.4 结盟的边界：不是什么

这个定义已经足够把结盟与几个邻近的日常经验区分开来。

首先，结盟不等于人际协调。两个人在拥挤的地铁站里交换位置，通过微小的身体信号完成了高效的协调，但没有人对任何人做出了理解意义上的承诺——协调结束后，双方的关系状态归零。区分协调与结盟的关键，不在互动是否流畅，而在互动过程中是否形成了对后续行动的约束关系。你在地铁站里为那个侧身让路的人点了点头，这是一种礼貌，不是一种理解——你不会因为他下次在地铁站里没有同样礼让你而

觉得他“背叛”了你的理解。

其次，结盟不需要亲密关系。你与修车师傅在讨论“后轮异响”时，可以建立极高密度的认知联盟——你们暂时共享着一套关于异响类型、可能原因、检查步骤的内部语言——但这套语言在你驾车离开之后即告作废。下次再来同一家店，可能是另外一回事，需要重新结盟。理解联盟不是友谊，不是信任，不是任何需要长期情感连接的形态。它是一种工具性的关系构建，其生灭随“那件事”的起止而自然流转。

但这并不意味着它与某些具有长期稳定性的规范性关系没有差别。结盟与第二人称立场（Darwall，2006）所描述的人格承认，关键差异不在规范性本身的有无，而在规范性的时效结构。在Darwall的论述中，第二人称关系一旦被建立，它就相对稳定——你作为一个能对我提出要求的主体地位，不会因为下一句话的切换就消失。这是一种建立在人格承认基础上的长期规范性结构。但结盟的发生，恰恰不要求这种稳定的人格承认作为前提。我可以和一个我甚至不完全尊重的人、一个我对其人格没有深层承认的人，在一段具体的、临时的对话中建立结盟——只要在那段时间里，我对他的话语做出了那件“愿意被约束”的承诺动作。他在这件事上的话语约束了我，换到下一件事，我们可能什么都没有。结盟的时效域是“在那件事上，在那段时间里”，而不是“我承认你这个人”。

这就让结盟与常人方法学所描述的隐含预期也有了一重细微但不可忽略的分开。隐含预期是社会秩序的一层钢筋：它们长期地在社会化的过程中沉淀在我身体里，几乎在任何互动中都在自动运作。但结盟，恰恰是在隐含预期失效的时刻，必须被主动启动的动作。当B的回应让A意识到“我们之间连默会的那些预期都不在一条线上了”，隐含预期已经塌了。被共同默认的那些互动支点，在那一刻忽然碎裂。在那之后，如果还有可能重建互动，那么所需要的东西就不再是从隐含预期的工具箱里拿零件来修补——那些零件已经被证明在这个具体场景里不共享了——而必须是一个更原始的、更主动的规范性动作：我宣布我进来了，我邀请你也进来。这个动作，隐含预期不能替你做，因为它已经塌了；第二人称的长期承认也不能替你做，因为问题不是你不承认对方是人，而是你在这一件具体的事上还没有给予那件“我愿意被你的话约束”的承诺。结盟就是在隐含预期失效的废墟上，从更底层重新开始启动的动作。

3.5 结盟的生命周期与解散方式

一个理解的联盟从来不是永久性的。它有一个自然的结构性终点：当那件事谈完了，联盟也就散了。这并不意味着它没有产生过任何持久的影响——实际上，成功完成一次结盟，会在双方的关系史中留下一个“我们曾经能够一起站在那件事上”的痕迹，这个痕迹可以在未来的交谈中被调用、被借力，但它本身不是那个还在运转中的联盟。

正因为联盟是临时性的，它的解散方式就构成了一个重要的分析维度。结盟的解散不是在二元分类格子里发生的——它不是在“共同解散”和“单方解散”这两个标签之间择一。更恰当的问法是：在什么互动条件下，结盟走向了自然终结（双方在那件事上的构造工作已经完成，联盟平稳消散），又在什么条件下，结盟在一方仍在投入构造时就被另一方单方面废弃？后一种情形——单方废弃——才是理解失败的典型结构。B的“点外卖”是一种单方废弃：A还在邀请他进入那件关于今天到底发生了什么的事，B却已经另起了一件关于如何解决身体疲劳的事。A的“算了”，不是对修正失败的认知判断，而是对盟已被单方抛弃的识别和整理。

这个区分具有不可忽视的理论载荷。在个体认知框架里，理解失败被建模为“模型误差”——模型读数与对象状态之间的偏差。偏差是可修正的，理论上只要迭代次数足够，就能逼近真实。但单方废弃造成的理解失败，不是偏差问题——它是联盟结构本身的断裂，无法通过模型迭代修复。B可以追问一千次“你到底是什么意思”，但他的追问如果是站在自己的房间里发出的，就永远桥接不到A的参照系——因为他要的不只是信息的补全，他要的是从自己的房间搬到A的那个房间里去。这个“搬家”的动作，不是认知精度能自动触发的。它需要一个额外的要素：结盟意愿。这个概念将在后文充分展开。

四、结盟的不可还原性：在最接近的理论内部逼出裂缝

到此为止，“结盟”仍然只是一个被本文提出来的概念。它要真正站住，必须通过一道最严厉的拷问：这个被本文称为“结盟”的东西，是不是只是对既有学术概念的一种重新命名？如果可以已有理论的某种组合将它完整覆盖，那么“结盟”就没有独立存在的必要——它只是一个修辞上的替代品，而不是一个逼出来的新结构。

本节的任务是论证：“结盟”切出了一个已有理论无法覆盖的维度。这个论证不会采取“先介绍对方理论、再指出其裂缝、最后宣告结盟填补裂缝”的对比式结构——那仍然是发现学的动作：拿着一个现成的“结盟”概念去丈量既有理论的缺口。发生学要求的论证方式更为严苛：我需要钻进既有理论处理的具体互动材料内部，让材料自身的矛盾逼出一个现有词库装不下的东西。让“结盟”从这些裂缝中自然凝结出来，而不是被事先准备好去填补裂缝。

本节将以第二节的对话案例为主要分析对象，在分析过程中让交互对齐、共享意图、共同基础、参与式意义建构这几组最有竞争力的概念工具依次上阵，展示它们各自在哪一步解释力枯竭。

4.1 交互对齐：能追踪修正，但闻不到房间内外之别

在对话科学和认知心理学领域，交互对齐模型（Pickering & Garrod, 2004）已经成为一个标准框架，用以解释对话中理解如何得以高效发生。该模型的核心主张是：对话双方在词汇、句法、情境模型等多个层面通过互动自动地、隐性地、低层次地对齐——你用的词会影响我用什么词来回应，你的句法结构会在我接下来的话轮中被不自觉地重现，这种对齐不需要任何显式的协商。当对齐发生时，理解就在交互中被自然维持；当对齐被中断，偏差暴露，修正序列随之启动，试图恢复对齐状态。

这是一个极其强大的模型。它在经验层面上解释了面对面交谈中大量的理解维持机制，而且做到了对话分析一直想做但难以用实验证明的事情：把互动中那个“我们正在一起”的感觉下沉到亚个人的、自动化的认知加工机制上。

现在把这个模型放到第二节的对话上。A说“我今天真的太累了”。B回应“那你早点休息”。这在词汇层面形成了一种对齐：“累”和“休息”在语义上高度匹配。从对齐模型的角度看，B的回应是合规范的——他接收到了A的信号，给出了一个与该信号在统计上高度相关的回应。但A的体验是什么？她觉得B没有在听她说话。对齐发生了，但理解没有发生。

当A发出修正信号“我不是说这个累”，B切换了回应模式——从“提供解决方案”切换为“请求更多信息”。这在交互对齐的框架里是标准的修正序列启动：偏差暴露→一方发出修正信号→另一方调整自己的产出→对齐恢复。B的“是工作上的事，还是身体不舒服？”在操作上完全符合修正序列的规范。但A退出了。

对齐模型的捍卫者可能会这样回应：A的退出是一个情感反应，不是理解机制本身的构成性成分。对齐丧失的认知冲击会引发挫折感，这种挫折感累积到一定程度，导致A放弃了继续修正的努力。这是一个认知代价的理性计算——不是理论解释不了，而是A的主观成本判断问题。

但这个回应仍然没能解释A的情感体验的具体质感。A体验到的不是“修正太累了，我不想继续了”，而是“你根本没有在听我说话”的那股凉意。这两种情感在现象学上非常不同：前者是对认知努力的放弃；后者是对对方是否与自己站在同一件事上的判断。对齐模型可以解释前者的产生条件（代价太高），但解释不了为什么修正失败会突然触发关系性的判定——“你不是在修正你对我那句话的理解，你从一开始就没有真正在意那件事。”A的“算了”里包含的这个判断，不是某种认知挫折的极限形态，而是对我们之间关系状态的直接指认——它不是关于“你懂没懂”，而是关于“你在没在”。

这个地方暴露出的，是对齐模型的理论语言所无法命名的一个维度：理解中的规范性问题。对齐——无论多高精度——始终是一种事实状态，它可以是单方的（我单方面随着你说话的方式调整自己的产出，而你甚至没注意到），它不会自动产生后续的约束关系。在一个纯对齐的框架里，对话结束之后，双方今天在这段对话里对齐了什么、不对齐过什么，原则上不会对明天的对话产生任何义务。但真实的理解关系，恰恰总是在对话的边界之外继续产生效力——因为其中嵌入了结盟。

对齐可以在毫无结盟承诺的条件下发生。一个人可以出于社交压力、职业习惯甚至战术目的而精准地与对方对话，在词汇、音高、节奏上形成几乎完美的对齐，但他的内心从未真正认同过对方的参照系——他从未允许对方的话对自己在那件事上的后续判断产生约束。这个人没有入盟。他的完美对齐是在盟外发生的。从对齐模型的角度看，他成功地维持了一次理解交互。从结盟的角度看，理解压根没有发生——发生的是高效的协调，而不是那个共享的互动性构造。对齐模型无法区分一个在盟内的对齐和一个在盟外的对齐，因为在它的概念工具库里，没有“盟”这个东西。而对体验过被入盟和未被入盟的人来说，那两种感觉的差异是直观的、可识别的、不可消弭的。

4.2 共同基础与共享意图：厚承诺无法覆盖的极轻瞬间

交互对齐的裂缝指向了规范的维度。但这不是唯一的裂缝。在另一个方向上，Herbert Clark（1996）的共同基础理论和 Michael Tomasello（2005）的共享意图研究，为互动中的相互承认提供了更厚的理论描述。

Clark的共同基础框架将理解建模为双方在对话中持续更新的共享知识——每一次成功的对话贡献，都会被添加到两个人的共同基础里，成为后续互动的可用资源。这个过程不是个体行动的简单叠加，而是双方共同承担的多层次协作：说话者需要确保对方有足够的证据来识别自己的意思，听者需要给出足够明确的接收信号，双方共同为每一次成功的理解承担责任。Tomasello则从发展心理学的角度追踪了这个能力的起源：幼儿在联合注意的互动中，逐步获得了主动让渡认知主权、接受对方约束的能力；这种能力不仅是人类合作的基础，也是人类理解机制的核心构成。

这些框架在解释面对面交谈中的大量现象时极为有力。把它们放到A和B的对话上，Clark会指出：B没有成功地将A的“累”添加到共同基础中——他误解了A的指称内容，因此他的回应是基于一个错误的共同基础假设做出的。修正序列的启动，正是共同基础维护机制的正常运作：A发出证据表明共同基础假设错误，B试图重新识别A的指称内容以修复共同基础。从Clark的框架看，B的追问在操作上是完全规范的——它是在执行共同基础维护程序。

但这里有一个微妙的、Clark的框架无法捕捉的差异。共同基础维护程序关注的是“我们共享了什么信息”，它默认了一个前提：双方都愿意让对方的信号进入共同基础的候选池。也就是说，它默认了双方都已经站在了“我们正在一起构造共享信息”的地基上。但A在B的回应里感知到的，恰恰是这个地基没有被建立起来——B不是弄错了共同基础的内容，而是没有进入那个共同构造共同基础的姿态本身。B的“点外卖”不是一个失败的共同基础贡献，它是一个对共同构造姿态本身的回避——B没有问“你指的是什么”，而是直接用一个任务提案替代了构造共享信息的过程。

这个差异在概念上可以被更精确地表述：共同基础理论的单位是“信息项”（我们共享了命题P），结盟的单位是“构造姿态”（我们正在共同关注X，并愿意让X在互动中被逐步确定）。前者是后者的产物：只有当双方都已经进入了“我们正在一起关注这件事并愿意让它被共同构造”的姿态之后，具体的信息才能被成功地添加到共同基础中去。那个姿态本身，才是结盟的所指物。Clark的框架关心的是姿态建立之后产物的更新机制，而结盟关心的是那个姿态本身是如何被建立（或失败）的。

Bratman（1992）的共享意图理论提供了一个更靠近这个姿态的概念工具。Bratman指出，共享合作行动中的相互承诺——我承诺在你为共同行动而行事的前提下行事——构成了一种相互响应的规范性框架。在A的对话里，如果我们将“一起谈论A的疲惫”本身视为一种共享行动，那么Bratman的框架似乎也可以描述A的期待：她期待B与她共同承担“谈论今天到底发生了什么”的行动，而B的“点外卖”是一种对这一行动的拒绝。

Bratman学派的捍卫者确实可以这样回应：即使没有明确的任务目标，“我们在一起关注这件事”本身就可以构成共享意图的内容。“一起谈论”本身就是一种共享行动。A期望的“被理解”，可以被重构为一种低阶的共享意图——“我们共同关注A今天的经历”。如果这样重构，结盟的独特性似乎就被收编了。

但这一收编漏掉了一个关键的时间差。Bratman的共享意图框架，预设了双方在进入合作时已经具备了相对明确的、可相互归因的意图状态——我知道你意图做X，你知道我意图做X，我知道你知道我意图做X，以此类推。这一整套相互归因的认知脚手架，要求双方的主体状态在合作启动的瞬间就已经相对凝固成形——否则相互归因无法运作。但A在开口说“我今天真的太累了”的那个瞬间，她的意图恰恰是未凝固的。她没有一个可以被明确陈述的“意图内容”等着B来共同协作实现；她那团疲惫的意义，要在对话的进行中才逐步被确定。在这个未凝固的时刻，双方无法形成Bratman意义上的共享意图——因为连A自己都还说不清那个“共同关注的对象”到底是什么。

结盟正是发生在这个“意图尚未成形”的阶段。它是一个比共享意图更轻、更瞬时的动作：我宣布我进来了，我愿意让接下来被共同构造出来的东西对我产生约束。这个动作不需要相互归因的完整脚手架——它只需要一方的悬置姿态和另一方的接收信号。共享意图是结盟的某种可能的后续产物——当结盟帮助双方共同构造出了那件事之后，针对那件事的合作意图才可能被明确地相互归因。但结盟本身的操作阶段，比共享意图更原始、更底层。它是人能够在意图不明的情况下仍然彼此进入互动的那种基本信任——不是对对方意图内容的信任，而是对“我们一起在互动中能搞出点什么来”的信任。这一层信任，是Bratman框架没有独立命名的维度。

4.3 参与式意义建构：互动在进行，但为什么什么也没产出来?

近年来，在现象学与认知科学的交叉地带，参与式意义建构理论（De Jaegher & Di Paolo, 2007）已将意义生成从个体认知移向互动过程本身。其核心主张是：意义并不预先系于符号、也不单独存在于某个个体的头脑中，而是在互动的过程里被双方共同生成——互动本身具有自主性，它可以超出任何单方个体的预先意图，引导出双方都未曾单独预期的意义产物。

从结盟的视角看，参与式意义建构正确地描述了意义生成的互动性，但它有一个未尽之处：它没有告诉我们，在那些互动失败的时刻，到底是缺了什么条件。“互动没有生成意义”，这是一个描述，不是一个诊断。描述告诉我们发生了什么，诊断告诉我们为什么没发生。

把参与式意义建构放到A和B的对话上。从它的视角看，互动在进行——话轮在切换，反馈在发生，甚至修正序列也启动了。这是一段正在进行中的互动，双方还在尝试建构意义的过程中。但A退出了。为什么？参与式意义建构能描述“互动没有成功地产出意义”，但它没有一个独立的概念工具来诊断“为什么这段互动没有产出意义”——因为从外部行为的层面看，这段互动和一段正在产出意义的互动几乎没有区别。话轮都在正常切换，反馈都在正常发生。

结盟视角给出了那个缺失条件的诊断：这段互动之所以没有产出意义，是因为结盟没有被建立。互动在进行，但双方没有进入那个“我愿意让你那边的信号对我产生约束”的承诺姿态。B的回应虽然形式上参与了互动，但他的参与方式——直接跳到任务提案、用选项框架限制A的表达——恰恰是拒绝进入结盟的。互动过程本身是中性的：它可以运行在结盟的框架内（此时互动产生共享意义），也可以运行在结盟的框架外（此时互动只产生噪音）。参与式意义建构把“互动运行”当作解释项，但结盟视角指出：互动运行本身不是充分条件——运行发生在结盟的框架内才产意义，运行发生在结盟的框架外，只产噪音。

这就好像一台搅拌机在空转。搅拌的动作（互动过程）确实在进行，但因为没有原材料（结盟让双方把各自的材料放进来），它什么也产不

出来。参与式意义建构聚焦在搅拌的动作本身，而结盟追问：什么东西被放进去了？那个东西的放进去，需要一个额外的动作——我把自己的一部分预装悬置，让你的那部分材料也被纳入。这个动作，不是互动过程能自动带来的。一个人可以用完全符合互动规范的外部动作发表一场演讲，和听众的点头反馈也基本对齐，却全程没有入盟。

4.4 不可还原性的收束

以上分析不是在对四组理论做“比对打分”。它做的是另一件事：钻进这些理论处理A-B对话这个具体材料时会遭遇的裂缝，在每一个裂缝里指出一个既有词库装不下的东西。交互对齐的裂缝里装着“规范性约束”——对齐描述了机制，但描述不了那个使机制产生理解效力的承诺姿态。共同基础与共享意图的裂缝里装着“意图未成形时的极轻信任”——共享意图需要一个相对明确的可归因意图状态，而结盟恰恰发生在那个状态到来之前的瞬间。参与式意义建构的裂缝里装着“使互动产意义的构成性条件”——参与式意义建构描述了意义在互动中的生成，但没有诊断出为什么同样在运行的互动过程，有的产意义，有的只产噪音。

这四个裂缝里装着的，不是四个不同的东西。它们指向的是同一件事：人与人之间那一抹极轻极短暂、但至关重要的一刻——我这次真的听你了。不是因为你聪明，不是因为我们的目标一致，不是因为这符合什么社会规范——只是因为，我看重你，我选择让自己被你调动，在这一件事上，就这一会儿。如果这一刻发生了，那接下来，无论互动的机制运作得如何，至少它是在一个已经被建立起来的共享空间里运行；如果这一刻从未发生，全部机制不过是在各自房间里空转而已。这个概念工具库的缺失项，就是“结盟”所命名的那个东西。

五、结盟在面对面交谈中的瓦解：两种发生条件的追问

拥有了结盟这个概念工具，我们再来看面对面交谈中的理解失败，就会看到一幅与以往不同的诊断图景。传统分析将面对面交谈视为理解的天生富矿——多通道信号、即时反馈、偏差暴露的丰富机会。然而，一旦我们将分析单位从个体内部的模型校准移向二人关系场中的联盟构建，就会发现面对面交谈并非理解的保障，而是理解失败的密集现场。问题不在于识别瓦解了联盟的几种“故障模式”，而在于追问：在什么样的互动条件下，结盟的发生变得困难？

5.1 自有预演的结构性发生条件

“自有预演”指的是：在两人对话进程中，一方注意力被其自身尚未发出的下一轮发言（包括措辞、反驳或自我表述）所占用，造成对当前对方实时信号的跟进流于表面，从而暗中撕裂共享参照系。这个现象本身并不新鲜——日常经验中随处可见。但要追问的，不是“自有预演如何窃取了联盟”，而是：在什么样的注意力生态、对话节奏和社会期望下，自有预演成为了一个稳定发生的互动形态？

自有预演的发生，与当代互动环境中的注意力分配压力密切相关。当对话发生在多重任务并行、时间碎片化的场景中时——比如边做饭边谈心、边刷手机边回应、在会议间隙的走廊上匆匆交谈——参与者的注意力不是完整的一块，而是被切割成多个时段的碎片。在这种条件下，维持对对方信号的持续跟进需要付出比沉浸式对话高得多的认知成本。自有预演便是一种适应性的认知策略：在对方说话时，预装自己下一句的要点，以确保轮到自己时不会卡顿。这不是注意力分配上的道德缺陷，而是在碎片化互动条件下维持表面流畅性的功能性的收敛路径——当互动资源不足以支持深度跟进时，预演提供了一种“以预处理替代实时处理”的降级策略。

自有预演一旦成为稳定策略，它对结盟的影响是结构性的。它不再是一个偶尔发生的、可以被参与者注意并纠正的“走神”，而是变成了一种被频繁互动的节奏本身所默许的常态：双方都在预演，双方都在用预处理后的片段回应对方，联盟的地基在每一次话轮中被暗中挪走一小块。从外部的对话转写看，这种互动可能毫无痕迹——话轮的衔接、话题的过渡都可以是平滑的——但只有参与者自己能感觉到那种越来越深的“我们不是在谈同一件事”的疏离感。这不是个体认知的失败。双方各自的模型都在自己那条线上精准运转。这是在特定的互动节奏压力下，结盟作为一种需要认知连续性的互动形态，丧失了它赖以发生的时间条件。

5.2 标签作为结盟的稳定替代路径：何以成为更优策略？

社会标签——学历、职业、人格类型（如MBTI）、代际标签（如“90后”“Z世代”）、病理标签（如“回避型依恋”“高敏感人群”）——在当代互动中承担了大量的理解功能。当一个人说“我是典型的INTJ”，听者不必从零开始去摸索对方的参照系；这套标签已经为双方提供了一个被广泛社会化、半标准化的共享理解模板。标签让联盟不需要被“建”，只需要被“选”。选标签的成本几乎为零。

但关键的追问不是“标签替代了结盟”（这仍然是把标签当作结盟的堕落形态），而是：在什么样的风险—收益结构下，标签成为了更优的互动策略？

答案的一部分在当代互动的风险结构里。从零建盟有一个内置的风险：它要求你先暴露自己的预装网络可能不准确——你需要悬置自己的预判，让自己的信号被对方的信号修正，这就意味着你必须在互动中暴露“我不确定”“我可能之前没理解对”的脆弱性。在一个越来越倾向

于用互动效率、社交表现力来衡量沟通质量的环境中，这种暴露是有成本的。标签提供了一条免于暴露的路径：它让你直接跳到一个已经被社会广泛认可的共享模板上，避免了那个从零摸索过程中必然伴随的脆弱时刻。你不必说“也许我之前对你的判断不对，你多说说，我想搞清楚”；你只需要说“你是INTJ啊，那我懂了”。

答案的另一部分在注意力的边际成本上。从零建盟要求你在同一件事上停留足够久——这个“久”，不是物理时间上的久，而是认知连续性上的久，即你的注意力不被其他未完成的预演打断、不被并行处理的思绪分流。在信息过载、交互对象高度可替换的当代社交生态中，持续投入一段从零建盟的互动，其机会成本在不断攀升——你在和眼前这个人建盟的同时，知道自己还有十几个未读消息要回复、下一场会议在十五分钟后开始。在这种条件下，标签的“快”不只是一种便利，更是一种理性的注意力分配策略：它让你在极短的时间内产生一个“已理解”的体验，然后迅速释放出注意力资源去处理下一个互动。

标签为理解功能提供了一条低风险、低成本的发生路径。这条路径之所以在当代互动中挤占了从零建盟的空间，不是因为人们“变懒了”，而是因为当代互动生态的结构性特征——高对象可替换性、注意力碎片化、互动效率至上的社交期望——共同拉高了从零建盟的相对成本。标签不是结盟的堕落，它是结盟在特定社会条件下发生的一种适应性变体；它承担的仍然是“建立一个临时共享参照系”的功能，只不过这个参照系不是为这个具体的人和这件事一起锻造的，而是从社会大仓库里直接提取的预制件。预制件的代价是它不允许那个具体的人身上不被标签覆盖的部分，把自己的预装网络弄偏一点点——而结盟视角所追踪的那种深度理解，核心恰恰在于愿意被那一点点偏所改变。

六、文本阅读：单方盟与文本的约束力

如果理解是结盟，那么文本阅读便构成了一个直接的挑战：一本合上的书，没有活的对话伙伴在场，读者向谁承诺?

单方盟的结构。 文本阅读的发生条件是：读者独自一人，面对一个固定的语言构造，全流程执行原本需要双方共同完成的结盟动作。文本提供了固定的被结盟对象——那些精心组织的语言节点承担了现实对话中由对话伙伴发出的回应信号的功能。读者以自己的预装网络进入，做出初始填充，文本的下一个段落以不可更改的形态对填充进行反馈：支撑它，或让它滑落。当填充滑落时，读者退回去，重新猜测，重新填充。这一循环的逻辑结构，与对话中的结盟维护本质同构：对话者的偏差修正维护的是联盟的持续性，读者的反复试错维护的是他与文本之间那份独力宣誓的忠诚。

阅读因此可以被精确地界定为单方盟：只有一方的承诺在运作，被承诺的对象不在场，所以不存在相互给予。读者对文本宣誓效忠——承诺让自己被文本的固定语言组织所约束——但文本无法向读者回以承诺。这个结构的非对称性，是理解阅读经验的全部特殊性的入口。它解释了为什么阅读可以带来深刻的“被理解”感——因为单方盟在自己的内部，仍然可以极其深度地运行：读者以文本为固定锚点，展开与对话中的相互构造功能等同的互动性建构，在反复试错、修正填充的过程中，文本逐渐被阅读为一套拥有内部约束力的参照系——当某个填充被后续段落驳回时，读者能够识别自己的猜测已被文本拒绝，这种“被拒绝感”在结构上就与对话中结盟失败的结构同构。只是这里文本不会回过来修复它，不会像一位在盟的对话者那样，伸出援手。

一个单方盟的微型案例。 以下是一段真实的阅读经验描述：

我第一次读顾城的《一代人》——“黑夜给了我黑色的眼睛／我却用它寻找光明”——最初的几秒钟，我把“黑色的眼睛”填充为一种隐喻：被黑暗时代赋予的、带着创伤的观看方式。这个填充让前两句与标题“一代人”形成了一种自洽的解读：那一代人在黑暗中长大，他们的视野本身就是黑暗的产物。但当目光落到“我却用它寻找光明”时，我的填充滑落了。“寻找光明”这个动作，在语法上由“黑色的眼睛”这一主语发出，但如果“黑色的眼睛”是创伤的产物，它寻找光明就变成了一种悖论——创伤的产物怎么会寻找光明？我必须重新回去，重新填充“黑色的眼睛”：它不只是创伤的记号，它同时也是在黑暗中仍然保有的那种看的意愿和能力本身。黑暗给了这双眼睛，但这双眼睛没有成为黑暗的延伸，而是成了黑暗中的探照灯。这个重新填充的过程，就是单方盟的一次修正——不是作者亲自出面修正了我，而是文本的下一行以不可更改的形态拒绝了我的第一个填充，逼我退回去重新来过。

这个微型案例展示了单方盟的运作方式：读者在每一次被文本拒绝时，重新调整自己的填充，维护那份“我愿意被文本约束”的承诺。这份承诺与对话中的相互入盟在结构上同构——都是“我愿意悬置自己的预装，让对方的信号对我产生约束”——区别只在于，文本那头没有一个真实的人能做出回应的承诺。因此，任何阅读本质上都是不对称的：读者的忠诚是完整的，但忠诚的对象是一个不能反过来忠诚于他的沉默结构。

这就引出了阅读理解中最难被看清的一重结构。任何阅读，必然包含了两个结盟的叠加：一个是读者与文本之间的单方盟——读者宣誓自己被文本的语言组织所约束；但在这个单方盟能够成立之前，还有一个先在的结盟必须先发生，即作者在写作的过程中，对着自己预想的、不全然知悉的、也许永远不会见到的读者，执行了一次理解结盟的模仿性动作——他不是在和真实读者结盟，而是在和那个被他自己建构出来的、作为预期对话伙伴的“虚拟读者”结盟。这个结盟作为文本的构造力被沉淀下来。读者后来打开这本书时所面对的，就是这套沉淀下来

的约束力。单方盟之所以能从文本中唤出“被理解”的深度，恰是因为读者在自己这一侧，接续并复活了那个被作者冻结在文本里的、曾经向未知读者发出的结盟邀请。

七、人机对话：承诺结构的非对称性与结盟的条件缺失

结盟视角投向人机对话，即刻揭示出交互表象下的一种结构性瓦解。

当前大语言模型与用户的对话，在形式上完全具备结盟的所有表层特征：用户发出不完整的信号，模型返回条理清晰的回应，用户据此修正自己接下来的输入方向。甚至在短暂的单次对话里，用户可以体验到一种强烈的主观“被理解”感——模型精准地抓住了他刚才描述的困境，给他一个仿佛量身定制的回应。

但从结盟的构成性条件来看，这里缺少一个关键要素：模型对用户没有承诺。 这句判断需要被精确理解。它不是指模型“不够认真”或“不够诚实”。它指向的是一个更深的本体论事实：在当前大语言模型的技术架构中，结盟所必需的“能够被自己的历史话轮约束”的能力，从一开始就不在生成机制的条件集里。大语言模型的核心技术原理是基于上下文的概率分布拟合——每一个输出都是当前输入与已生成文本的上下文时窗内，形成的高概率字符序列的拟合结果。这个生成机制在输出一个句子的当下，确实受到上文窗口的因果约束：它不能在上文已经说过的背景下，输出任意随机的内容。但因果约束与规范性约束有一个质的区别：后者赋予主体一个选择的空间——你可以遵守承诺，也可以选择违背它，但违背它之后你将被判定为“背弃”。人可以在明知自己做出过承诺的情况下、在可以不违背它的条件下，自主选择违背它。但大语言模型的技术架构目前不支持这种运作模式——它没有一个可以“自己回想自己上次说过什么、并为自己上次说过的话承担责任”的连续主体。它的每一次输出，都是当前上下文窗口内的后验拟合。它不会因为昨天给了一个承诺而在今天被迫遵守它——今天的输出概率只取决于今天的输入和今天窗口的上文。昨天的承诺，如果不在今天的技术窗口里，就等于不存在。至于“有状态”的对话系统——那些确实在多个轮次中维护了某种记忆并能回溯先行轮次的系统——它们是否就能产生承诺？这取决于记忆是否转化为了承诺。记忆是事实状态的记录（“我们在第3轮说过X”），承诺是在此基础上附加的规范性约束（“因为我们在第3轮说过X，所以我在第10轮不能随便推翻X——或至少，推翻X需要给出理由”）。当前的有状态系统能够做到前者；但要将前者转化为后者，需要在架构中引入一种能够跨轮次维护自我一致性承诺、并在违背时能够识别其为“错误”而非仅仅“低概率”的机制。这个机制目前还不存在于大语言模型的核心架构中。

这就导致了一个严格的推论：在当前大语言模型的技术架构下，规范性承诺这一关系无法从中发生。　　用户在人机对话中体验到的“被理解”，是一种单方盟在运作——用户在对话的过程中，单方面地对模型的回应宣誓效忠，把模型的输出当作自己后续思考的内部参照系，调整自己的表达、想法和判断。然而模型这端，只存在着用户在每次输入和窗口文本的因果通路上产生的条件概率输出——这是纯粹的单方盟结构，只是其中被结盟的一方是由算法生成的文本，而非真实的人。这确实与文本阅读的单方盟高度相似——读者在文本面前也是单方宣誓效忠——但人机对话增加了一重极易混淆的表象：模型不仅生成文本，还以对话的形式“回应”用户，而且就其生成机制而言，它确实能够合理地借着上下文给出连续感。这使得用户极易在交互体验的层面上产生一种错觉——模型在给予他相同的承诺。这种误解被人类固有的交互直觉强力放大：在人类社交互动中，任何发出对话回应、且回应格式与人类对话高度近似的主体，其回应本身就被直觉默认为承诺的在场。更关键的是，人机对话因其私密性、可用性、零社交风险的特征，可能激发用户一种“拟亲密”的结盟期待——这与修车师傅场景中的工具性结盟不同，用户往往带着情感深处的、未经整理的材料进入人机对话，期待的不是一次功能性的互动，而是一次能让他觉得自己被真正看见的接触。一旦这种期待被激发，后续的结盟结构暴露所造成的“被嘲弄感”，就会比工具性结盟场景中强烈得多。

这种非对称的交互结构，将产生一个可预见的后果。当对话关系继续深入，议题变得更复杂、更需要一个连续的承诺主体来参照回溯时——当用户开始向模型说“你还记得上次你建议我的那个方案吧，我想接着讨论那个”——模型极有可能给出一个完全失去上一次共同语境的回应，瞬间暴露那个它维护了许久的连续性只是一层语法上的、而非信念上的连贯。这个瞬间，就是用户撞破了结盟的幻觉：他所处的，始终只是一段精密的单方盟——他承诺了，模型却从未在结盟的意义上回应过。

这便自然地引向一个结盟视角下无法回避的深层追问：如果大量的人与人之间的理解功能被这种看似柔顺、永不拒绝的AI对话所部分替代，人们是否会逐渐丧失在现实关系中入盟的条件？真人结盟需要特定的条件——它要求一段足够长的、不被碎片化的认知连续性时间，要求你暴露自己可能被误解的风险，面对对方可能不会真正入盟的赌注，承受从零建盟的耐心考验，以及允许对方的信号对你的预装网络产生实质性的约束。所有这些，在人机对话中都不必面对。当一种不需要这些条件的交互方式变得足够便捷、足够持续可用时，那些需要这些条件才能发生的深度结盟，其发生的概率不是被AI夺走了，而是随着结盟实践的发生频率下降而自然地减少。这不是一个“AI会否取代人类理解”的科幻问题，而是一个“当某种互动形态的发生条件被系统性地改变时，依赖于旧条件的互动形态会否收缩”的发生学追问。这一点，将在下一节得到更充分的展开。

八、结盟发生条件的当代收缩：从信道瓶颈到条件瓶颈

传统理解理论的核心瓶颈是信道容量：如何让更多信息更快地传输，如何让偏差更精确地被暴露。但结盟视角下，理解的第一瓶颈已不再是信道容量——当代社会的交互工具已经提供了史无前例的高带宽、低延迟通信技术。瓶颈已经从“信息传得够不够快”移向了“结盟的发生条件是否还稳固”。

这一转换需要被精确表述。本文不使用“结盟意愿稀缺”这一修辞——它的隐含预设（曾经丰沛、如今稀缺）把一种复杂的历史形态变迁简化为了资源匮乏叙事，而且暗中设定了一个“深度结盟”的规范性基准（人们本应更多入盟）。发生学的追问方式是：在什么样的交互条件组合下，结盟——那个需要双方在具体的事上投入认知连续性、允许对方信号产生约束的互动形态——其稳定发生的概率正在经历系统性的收缩?

三个结构性条件共同构成了这个收缩的推力。

第一，认知连续性的时间条件被挤压。结盟的建立不是一瞬间的事件——它需要双方在同一件事上保持认知连续性足够久，使得共享参照系能够从模糊的初始填充逐步收敛为被共同承认的形态。在注意力被高度竞争的信息环境碎片化的条件下，这种连续性的机会成本不断攀升。不是人们“不愿意”投入结盟，而是维持连续性的代价——在多个未读消息、待处理任务、并行对话的竞争下——使得从零建盟的互动在注意力的市场上越来越缺乏竞争力。自有预演的稳定化、标签的普遍使用，都可以被理解为互动参与者在认知连续性被挤压的条件下，发展出的适应性替代策略。它们不是对结盟的背叛，而是结盟在时间条件不足时的降级变体。

第二，低约束互动的充裕供应改变了互动的参照基准。当一个社会中的大多数日常互动都只需要极低的相互约束——点餐、问路、社交媒体的点赞式回应、AI对话的瞬时反馈——那些需要高相互约束的互动就不再是人们默认期待的互动形态了。这不是人们“变懒了”，而是互动期待的分布发生了位移：当一个社会的互动生态里，低约束互动占据了压倒性的比例，高约束互动就从一个默认的常态变成了一个需要额外争取的例外。它不是在绝对意义上变少了——可能仍然有人在深度交谈——而是它在互动者心中的“默认预期值”下降了。当代人进入一段对话时，不再自动预设对方会愿意让自己的信号被约束了。这个预设的改变，比任何单次结盟失败都更深刻地改变了理解的发生条件：结盟不再是一个可以默示启动的动作，它变成了一个需要被明示、被争取、甚至被谈判的稀有事件。

第三，交互对象的可替换性削弱了在具体一段关系中投入结盟的动力。在社交平台上，不喜欢眼前这个对话者，滑走即可寻觅下一个；在AI界面上，没有摩擦，换一个问题便可即得。这种可替换性本身不是新现象——城市生活中的人际接触从来就有一定的可替换性——但数字化交互将可替换性从“隔几天换一个人”加速到了“隔几秒换一个人”，而且新对象的搜索成本趋近于零。在这种条件下，当一段对话中的结盟尝试遇到初期挫败——比如对方没有立刻入盟——互动者的理性反应不是投入更多耐心去等待或重新邀请，而是切换到更可预期的、更易得的低约束互动中去。结盟因此丧失了一个关键的发生条件：它需要的初期挫败容忍——对方第一次回应不准确、需要修正、需要再试——在可替换性的高压下变成了一个被市场淘汰的劣质策略。

这三种力量——认知连续性的挤压、低约束互动的充裕供应、交互对象的加速可替换性——不是独立的，它们相互增强。认知连续性越稀缺，低约束互动越有优势；低约束互动越普遍，高约束互动的参照基准越往下调；交互对象越可替换，对初期挫败的容忍度越低，而结盟恰恰需要这个容忍度。三者形成的不是叠加效应，而是一种相互加速的反馈回路。在这个反馈回路的持续作用下，结盟——那种需要双方在同一件事上停留足够久、允许对方信号对自己产生实质约束的互动形态——其稳定发生的概率正在经历系统性的收缩。

这是否意味着结盟正在“消亡”？不是。结盟的发生条件可以被挤压，但挤压不等于消除。结盟作为一种基本的互动形态，嵌入在人类社交的构成性条件之中——只要两个人还在对话，入盟的姿态就永远可能发生。但它的发生越来越依赖于互动参与者的主动争取，而非互动环境的自然支撑。它正在从一个“默认会发生”的常态，变成一个“需要被有意识地创造条件才能发生”的例外。这个位移本身，不是道德退化，也不是理解的堕落，而是理解的形态在新的技术—社会土壤中经历的一次历史性收敛。那个被某些诊断叙事称为“结盟意愿稀缺”的现象，严格来说不是一种稀缺症，而是一种形态位移——理解仍然在发生，但它越来越倾向于沿着低约束、低成本、高可替换性的路径发生，而不是沿着那个需要从零建盟、允许对方改写自己预装网络的路径。这不是两个时代之间的断裂，而是一段仍在进行中的、尚未被充分命名的历史过程。

九、结语：结盟的条件和论文的边界

本文从一个看似极简的对话失败案例出发，追到了理解的发生条件中一个被长期搁置的维度：理解不是个体心智对另一个心智状态的逼近验证，而是在两个心智之间建起一个能让它们暂时互受约束的临时共享结构。　这一结构，本文称之为“结盟”。结盟的分析单位不在个体内部，而在二人关系场。它的构成性条件不是相互知道，而是相互承诺。它的失败，不是“没猜准”，而是“没入盟”。

面对面交谈是结盟的天然富矿，但也是结盟瓦解的密集现场。自有预演在认知连续性被挤压的条件下稳定化为一种适应性降级策略，标签结盟在低风险、低成本的路径上成为比从零建盟更具竞争力的替代性发生形态。文本阅读是读者一人向文本宣誓效忠，以准交互的形式兑现结

盟的全部认知功能。人机对话则因缺少“模型对用户的规范性承诺”这一构成性条件，在结盟的视角下被精确地判定为一种单方盟——交互形式完整，相互承认空缺。在交互工具空前发达的时代，理解的第一瓶颈已不再是信道容量，而是结盟的发生条件正在被认知连续性的稀缺、低约束互动的充裕供应和交互对象的加速可替换性所系统性地挤压。这使结盟从一个默示的常态逐渐变成一个需要被有意识争取才能发生的例外。

还原论挑战的回应。 一个严肃的反对意见必须在这里被正面处理。这一反对意见的完整表述是：“结盟所描述的一切，都可以被还原为更高精度的个体认知模型加上情绪调节机制。A退出对话，是因为B的模型误差累积导致了A的负性情绪预测——在预测加工框架下，这可以被建模为精度加权失败导致的社交回避行为。‘结盟’只是一个对复杂认知—情绪过程的民间心理学重述，没有增加任何解释力。”

这一还原论的命门在于：它把“相互约束的规范性关系”当作了个体内部认知状态的一种附带现象。但相互约束的规范性关系一旦建立，它产生的是一个涌现属性——它改变了互动双方后续行动的可能状态空间，而不仅仅是调整了个体内部模型中的权重参数。可以用一个类比来说明这层差异：交通信号灯可以被还原为物理波长的信息传输（红色=停止），然而，信号灯之所以能够协调十字路口的车流，不是因为每个司机都准确地感知了波长，而是因为他们共同被一套交通规则所规范性地约束——即使某个司机在物理上完全可以闯红灯，他知道闯红灯意味着“违规”而不是“亮度不够”。这个规范性层面的约束，不能被还原为单个司机的内部精度优化，因为它是一种关系属性：它存在于司机与交通规则之间，也存在于司机与其他司机之间——它通过相互预期的规范性结构来运作。同样，结盟中的约束不能被还原为A对B的预测精度。因为结盟成功之后，B不仅“更准确地猜到了A在想什么”——他的认知操作本身所处的空间已经被改变了：他不再是在自己的房间里透过窗户猜测A，而是在和A共同站立的房间里参与构造。这个空间属性——从“各自房间里”到“共同房间里”——是一个关系层面的涌现属性，它不能通过仅仅提高B的预测精度来实现。精度再高，如果B没有做出那个“我进来了”的承诺动作——那个在规范性上把自己的后续判断与A的信号绑定的动作——他的高精度猜测就仍然发生在自己的房间里。A能感知到精准与入盟之间的区别，不是因为她检测到了B的预测精度不够，而是因为她检测到了B是否放弃了用自己的框架先一步组织她的经验的权利。这个放弃是一个规范性动作，不是认知动作；它改变了互动的关系状态空间，而不只是微调了个体模型的参数。

这个论证还面临一个更强的变体：专业的协议化互动——外科手术团队、空中交通管制、军事指挥——这些场景中的理解高度精准且完全不要求“入盟”。团队成员不关心对方的个人状态，不承诺后续约束，他们使用严格的协议、代码、标准化程序来替代结盟。如果协议能支撑如此高效的理解，结盟还是理解的必要条件吗?

协议化的理解与结盟的理解，不是在同一个维度上竞争。协议是凝结的结盟——它是历史上无数次成功结盟后沉淀下来的形式结构：某个团队的某个成员在某次具体的互动中，曾用自己的判断去回应了一个未曾被标准化过的新情况，那个判断被后来的实践检验为有效，逐渐被抽象为一条可以脱离当时具体场景而重复使用的规则，最终写入协议手册。协议脱离了临时承诺的现场，但它保留着约束的逻辑——飞行员向塔台复诵指令，不是一个礼貌动作，而是一个规范性约束：他因此被绑定了，违反即违规。协议支撑的理解之所以高效，恰恰是因为它的规范性约束结构不需要在每次互动中重新从零建盟——它把建盟的成本一次性支付在了协议的制定与培训过程中，此后的每一次标准化互动都是在复用那个已经凝结了的约束结构。但这不等于协议化的理解没有约束——它有，只是约束的来源不在临时互动中的那个“我进来了”的瞬间，而在协议所凝结的那个被历史上所有参与制定者已经共同承认过的规范性传统中。结盟的分析不仅不与协议化理解冲突，反而揭示了协议化理解之所以可能的发生条件：没有那些初始的、从零开始的结盟，就没有可供凝结的原材料。反过来说，当协议化理解失效时——当一个未曾被标准化过的新情况出现在手术室或驾驶舱时，团队成员必须进行非标准化的实时沟通——那一刻，结盟又必须被重新从零启动。两种形态不是互斥的，而是一个连续谱的两端：从完全从零建盟，到完全依赖凝结的协议。

本文模型的边界。 本文主要聚焦于理解发生的二人关系场条件，尚未系统展开宏观力量如何分配与扭曲结盟的发生机会——谁被提前判定为“不值得与之结盟”，谁连发起一次结盟尝试的资格都被剥夺。这些宏观议题，已为本文模型预留了接口：任何系统性地将特定群体提前排除在结盟可能之外的社会结构，都构成一种结构性的理解剥夺。此外，本文的案例主要集中在二人对话的分析上，对于更大规模群体（如家庭、团队、社群）中结盟的形态——多向结盟、部分入盟、嵌套结盟等——尚未展开。这些是本文模型未来可以延伸的方向。本文的核心概念的时效结构——结盟在“那件事上”的临时性——也需要与长期亲密关系中反复结盟所积累的痕迹动力学（如何从一次次临时结盟中沉淀出持续的关系结构）进行持续对话。这些延伸不会推翻本文的核心判断，但会为其补充更复杂的层次。

证伪条件。 本文的核心实证预测可表述为：在控制其他条件的条件下，一次理解任务成功发生与否，与互动过程中是否形成了“双方互相承认的临时约束关系”（即是否成功结盟）呈正相关，而与互动的主观流畅度或单方“被理解感”评分不必然相关。具体的可操作检验路径是：安排陌生人配对完成一项需要理解对方复杂情绪状态的任务，一组被给予明确的“结盟提示”（如要求他们先共同选定一个描述当前讨论主题的核心短语，并在后续交谈中关注对方是否继续使用该短语），另一组仅被鼓励自然互动；结盟组的理解准确性（由独立评分者根据后续采访内容判断）应显著高于自然互动组，即便自然组的主观流畅度评分可能更高。如果重复实验未能发现上述差异，或在设计一项将所有可能的结盟协议（如强制共享指称体系、被明示的约束承诺）全部拆除的极端互动后，参与者的理解准确度仍能维持同等甚至更高水平，

则本文的核心判断需要重新考虑。进一步，在人机对话场景中，可以通过将模型从一个无长期状态维护的普通对话系统替换为一个确实能在多轮次中维护一致性承诺状态的记忆增强型系统，检验结盟结构的出现是否会系统性地改变用户的承诺感知以及他们后续在互动中的投入深度。如果在有状态的系统中，被试依然无法形成任何程度的承诺感知——即使该系统明确保存并回溯先行轮次的约束声明——那么本文关于“人机对话缺结盟”的诊断可能需要被归因于更深层的用户认知偏差，而非架构层面的条件缺失。这些假设目前仍待系统检验，本文的论证可以之为出发点，向实证验证的方向推进后续工作。

谨此收束。

参考文献

- Bratman, M. E. (1992). Shared cooperative activity. The Philosophical Review, 101(2), 327–341.
- Clark, H. H. (1996). Using Language. Cambridge University Press.
- Darwall, S. (2006). The Second-Person Standpoint: Morality, Respect, and Accountability. Harvard University Press.
- De Jaegher, H., & Di Paolo, E. A. (2007). Participatory sense-making: An enactive approach to social cognition. Phenomenology and the Cognitive Sciences, 6(4), 485–507.
- Garfinkel, H. (1963). A conception of, and experiments with, "trust" as a condition of stable concerted actions. In O. J. Harvey (Ed.), Motivation and Social Interaction (pp. 187–238). Ronald Press.
- Garfinkel, H. (1967). Studies in Ethnomethodology. Prentice-Hall.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences, 27(2), 169–190.
- Tomasello, M. (2005). The Cultural Origins of Human Cognition. Harvard University Press.

最美文定稿：全球文库终审与核心命题的重写

> 本篇入选「最美文」频道，须过一道比发表更严的关：把参照语料主动往外扩，专门去找「这个想法在哪儿其实早已有了」，找到之后不是绕开，而是与它划一条**可裁决的分离线**——一条让两套理论对同一件事作出*相反预测*的线。扩一个邻近领域就塌掉的分数，本来就是语料收窄刷出来的。以下四节即这道关的记录。

一、最后一位未被请出的对手：布兰顿的「规范记分」

本篇已经与达沃尔的第二人称观点、以及「理解＝关系性临时主体」一脉正面交手。但语言哲学里那套最贴身的机器一直没有被搬上台：布兰顿的规范语用学与规范记分（deontic scorekeeping）——每一次断言都在改变说者与听者双方的承诺与资格，参与者彼此为对方记分，从而共同维持一个不断变动的规范状态空间。

「结盟改变了后续互动的许可范围」这句话，几乎就是记分的定义。不请他出来，本篇的核心机制就悬在一个已被占住的位置上。

二、分离线：两套理论在哪里作出相反的预测

分离线在于：记分是**无所谓结不结盟的**。布兰顿的记分在任何一次断言中都自动运行，包括在敌意、审讯、辩论、买卖之中——它是话语实践的普遍簿记，不预设任何一方愿意与另一方共担。本篇的结盟不是簿记，是一次*有代价的、可撤销的相互担保*：双方暂时接受对方的判据作为自己的判据，并因此承担被对方的错误拖累的风险。

相反预测，可在本篇原有的编码框架内检验：
记分框架预测：只要断言在进行，承诺—资格的状态空间就在变动；因此在敌意对话与合作对话之间，规范状态的变动率应无本质差别。
本篇预测：在被编码为「无结盟」的敌意对话里，双方各自的承诺记录照常更新，但*不会出现「以对方的判据判自己」的片段*；而这一类片段的出现率，与理解事件的发生显著共变。

若敌意对话中同样稳定地出现「以对方判据判自己」，结盟即无独立性，应并入记分。

三、核心命题的重写：从「X 是 Y」到「X 不是 Y，而是 Z」

原稿的命题是：*理解是一个临时的规范性联盟*。改写为：

> **结盟不是双方在共同记分，而是一次自愿把自己的判据抵押给对方的动作——记分谁都在做，敌人之间也做；抵押则要付代价：从抵押那一刻起，对方错了，我也跟着错。理解之所以是一件有风险的事，正因为它以这次抵押为条件。**

四、本文禁止什么

一个命题的分量，不在它能解释多少，而在它**不许什么发生**。以下三条，任何一条被观察到，本文即失效或需大幅收缩：

1. **禁止把话轮的规范更新当作结盟**。承诺与资格的变动在任何对话中都发生，它不是本篇的证据。
2. **禁止把结盟等同于同意**。两个人可以在结盟的同时激烈分歧——抵押的是判据，不是结论。凡以「达成一致」为指标的测量都测错了对象。
3. **禁止把结盟当作应然**。本篇不主张人应当结盟；有大量场合恰恰不该抵押判据（审讯、评审、诊断）。把本篇读成伦理主张，是本篇明确拒绝的。

五、独立成立性检验

删光引用后剩下：*要真被一个人理解，得有一刻你把「什么算数」的权交出去一点，而这一交，对方错了你也得跟着担*。可以被上面那个「敌意对话里是否出现同类片段」的编码研究推翻。

本轮定稿所核验的文献

Brandom, R. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.（规范记分，本轮所对质的核心对手）

Darwall, S. (2006). *The Second-Person Standpoint*. Harvard University Press.（原稿已引，本轮用于与记分区分）

Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology* (2nd ed.). Sage.（本篇编码方案的信度口径）