Lieder 2019实验一相关细则

被试获得奖励是在各种不同条件下进行排名而获取奖励，排名标准指标是 24 个 trials 后统计的 real rewards 而不是 points，points 与实验奖励并不直接相关，points 只是一个暗示，影响被试的行为。但最终获取奖励取决于真实的 total real rewards。

指导语如下：

> When you submit the HIT, you will receive a bonus payment that reflects your performance relative to the performance of other players. The top 1 percent of the players will receive $2, the second percentile will receive $1.98 ,..., the 50th percentile will receive $1, ... , and the worst 1 percent will receive a bonus of 2 cents. Depending on which version of the game you will be assigned to losing points may be inevitable and winning points may be very difficult. Similarly, since the end of the game is determined at random, you will sometimes lose points even if you do the best thing possible. Don't worry about it! Your performance will be evaluated relative to the performance of other people playing the same version of the game. Hence, as long as you play as well as possible you will receive a high bonus even if you are constantly losing points. Conversely, winning points does not guarantee that you will receive a high bonus. To receive a high bonus you have to play better than the other people. Good luck!

被试的选择动机，它看到的线索，是线上的 points，这个 point 是由原始的 real rewards + pseudo rewards 共同形成。

但是在实验一的数据分析中，分析的是 real rewards 而不是被试获得的 points 值，也就是隐藏起来的 real rewards 值。

real rewards 在各被试条件下均由下图的真实 Map 给出。即无论在什么条件下，记录被试路径，以该路径回到下图中进行 total rewards 计算。

在各实验处理下，被试并不知道 real rewards 的值。

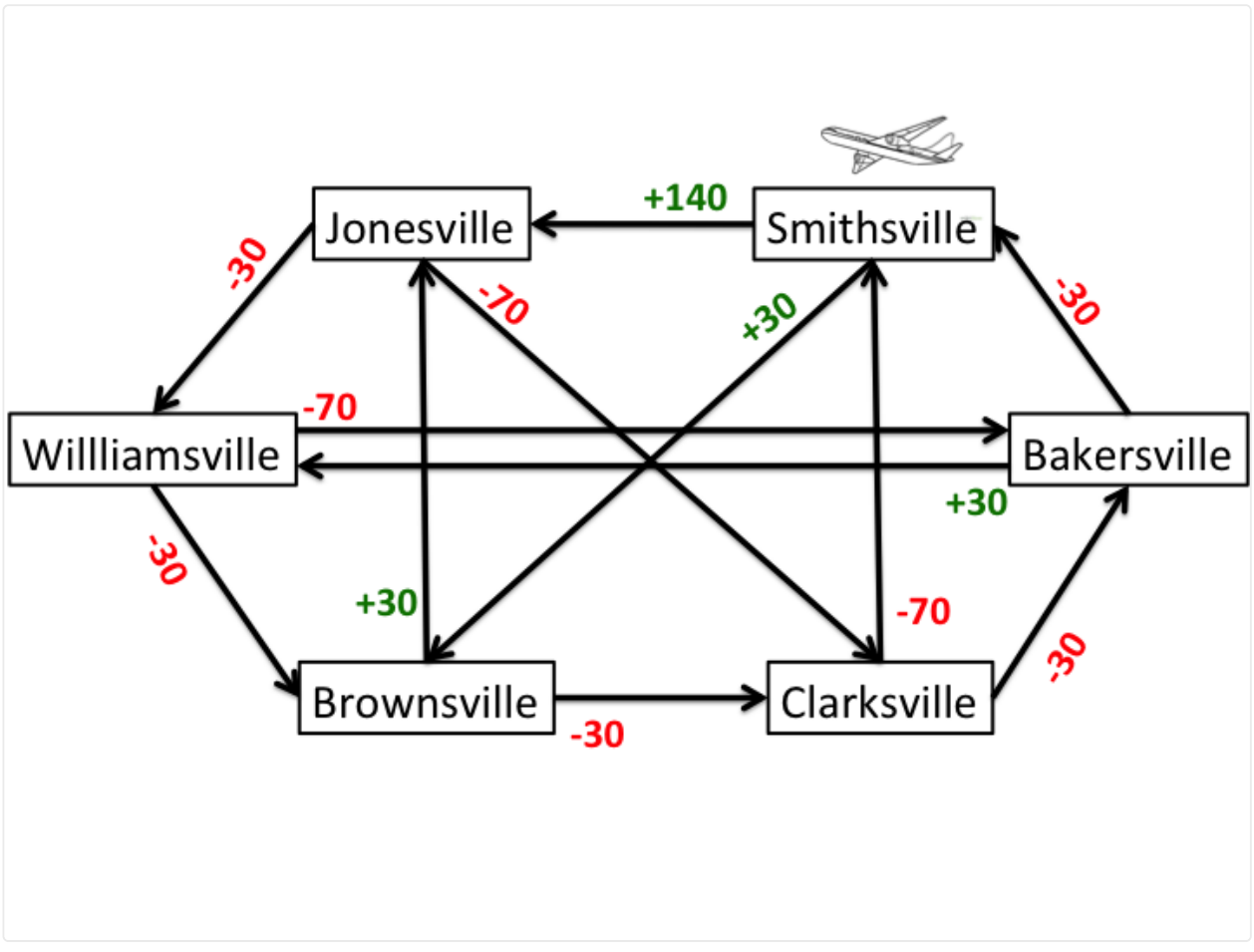

图 1 real rewards

第一个实验处理：Optimal Pseudo rewards

注意，这张图上线上的 points，是由图 1 中的 real rewards + 拟合出的 pseudo rewards 得到的。

在此处，被试在任何一个位置，只要选择 point 最大的路径，在 real rewards 中，就是最佳路径。

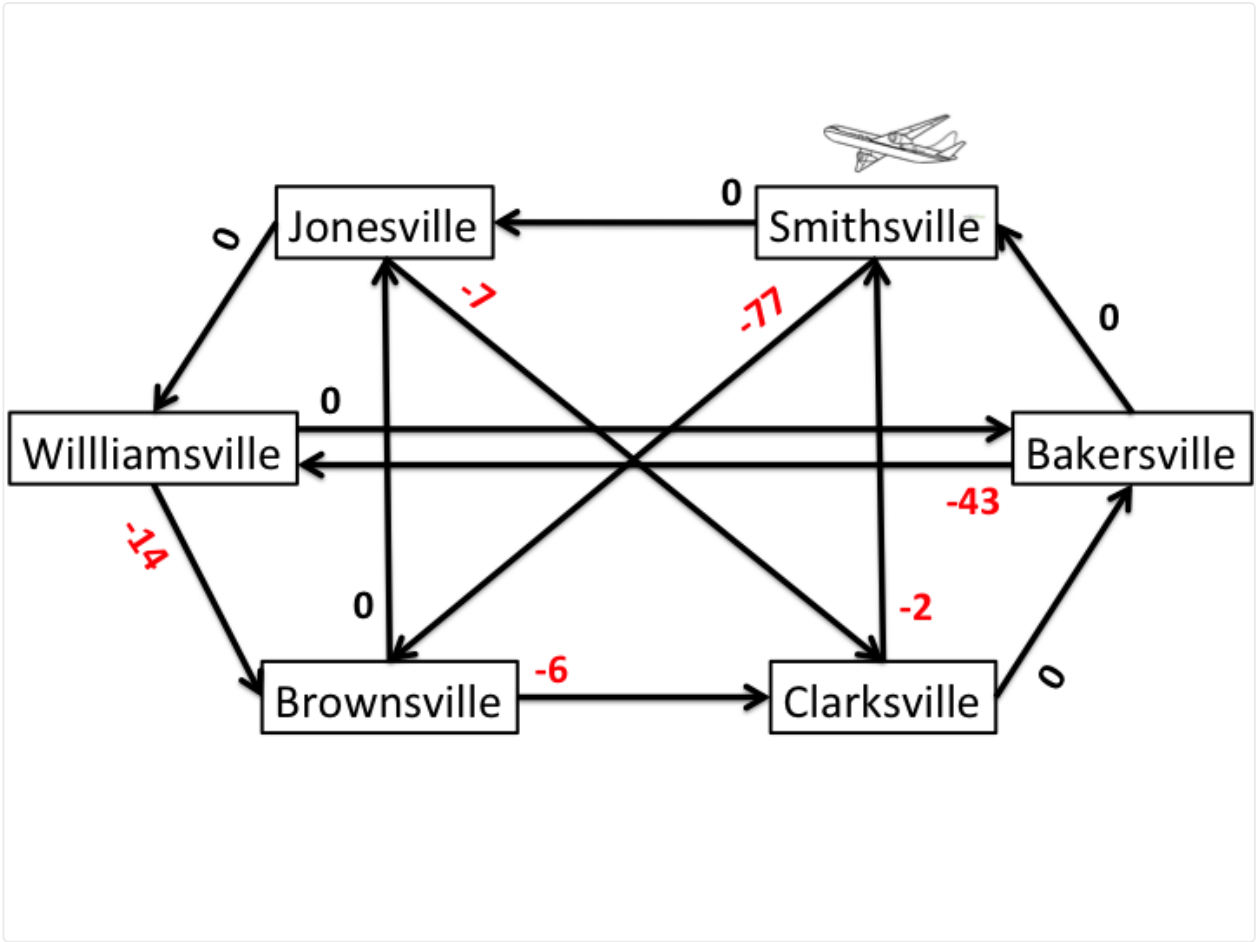

图 2 Optimal Pseudo rewards

第二个实验处理：Approximate Pseudo rewards

于是乎这个实验处理反而没有那么好理解了，不过可以简略推测出一些东西，在这个实验处理中，接近目标（Smithville）的 pseudo rewards 为正值，远离的为负值，如果距离不变，则 pseudo rewards 为 0。比如对比图 1 与图 3 Jonesville 到 Willliamsville，图 3 的 point 与图 1 的 real rewards 相等，侧面说明 pseudo rewards 为 0。

但是具体计算方法还是不清楚，比如对比图 1 图 3，可知 Brownsville 到 Clarksville 的 pseudo rewards 为 58，58 的计算方法还有待研究。

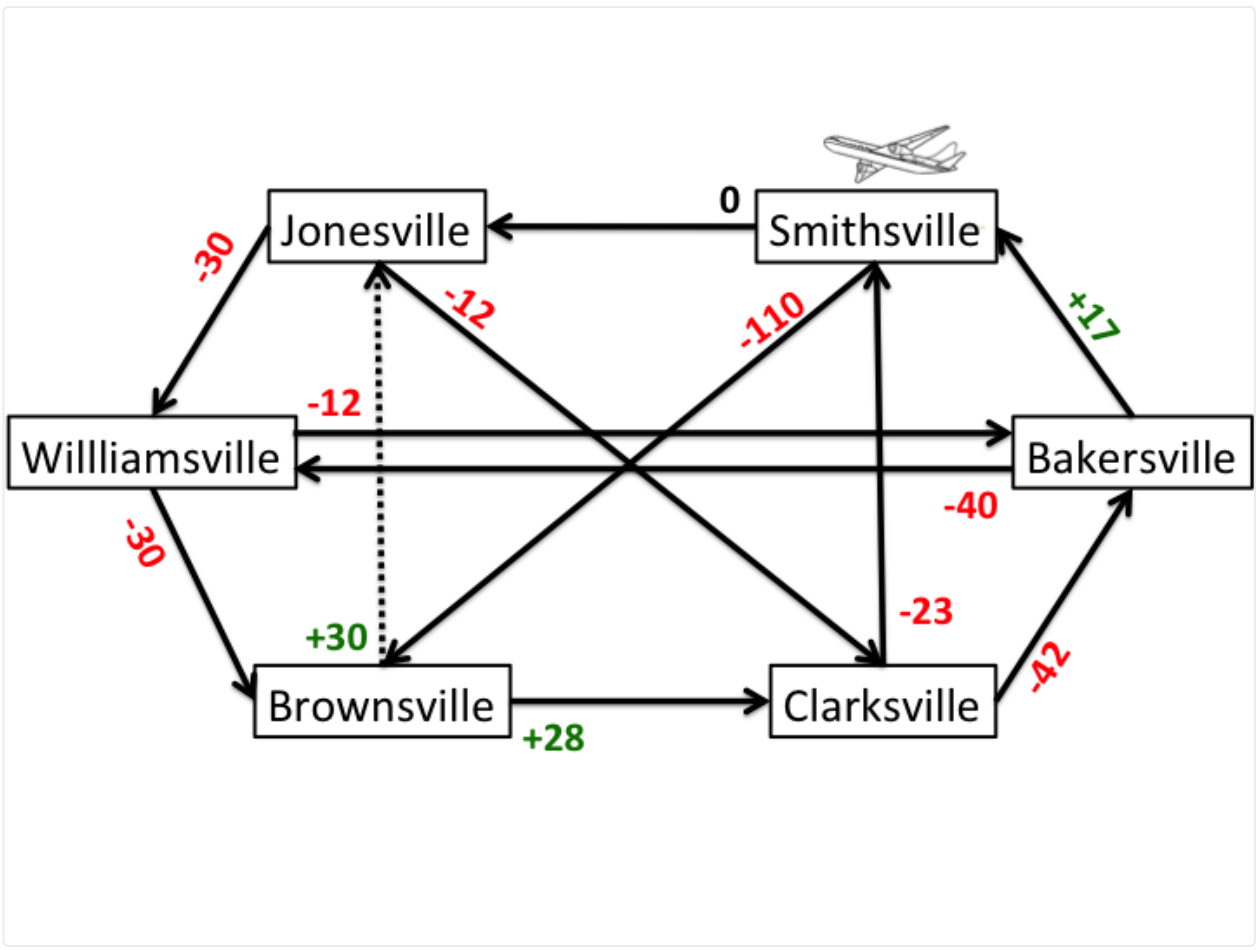

图 3 Approximate Pseudo rewards

第三个实验处理：Bad Pseudo rewards

这个实验在讨论时并没有研究清楚，但其实非常简单，是最简单容易获得的图形，比图 2 来源还要简单。

我们再仔细对比一下图 1 和图 4，原文中所谓 *The heuristic pseudo-reward was +50 for each transition that reduced the distance to the most valuable state (Smithsville)* 其实一看就明白，如果一个步骤移动后，当前飞机位置与 Smithsville 距离变近，那么我们的 pseudo rewards 就为 +50，远离或不变没有惩罚。

这个规则，对比一下图 1 与图 4 Jonesville 到 Clarksville 线上 points 的差异就能一目了然。

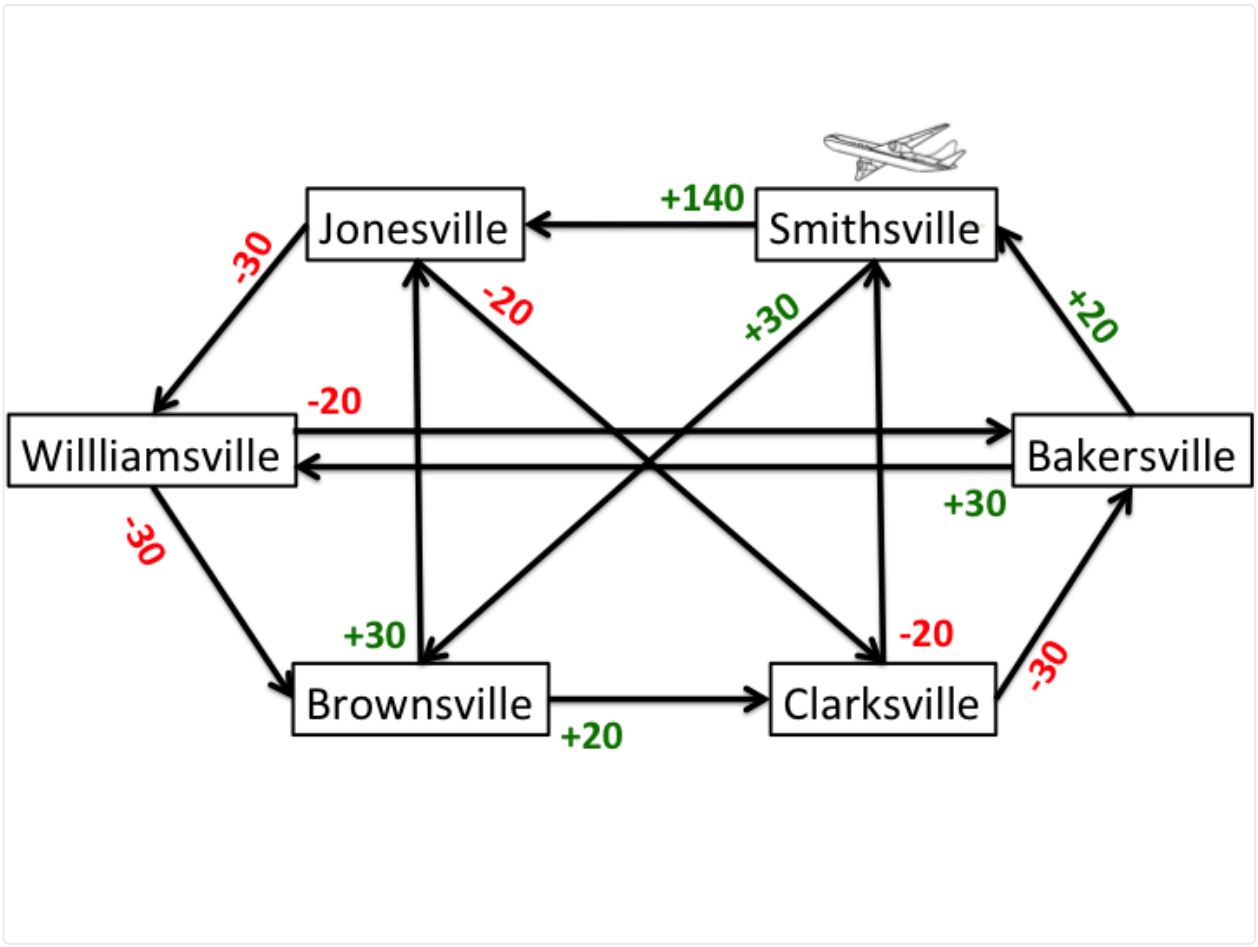

图 4 Bad Pseudo rewards

这样的实验设计，在两个选项中，会发现缩短距离的选项往往在当下会收获较高的 points，比如你现在处在 Jonesville，你有两个选择，去往 Williamsville 或者 Clarksville，选择前往 Clarksville 在当下的 points 收获是最高的，但是并不是总体 points 以及 real rewards 最优的选项。
这个实验设计其实是非常有意义的，它相比与第一个实验设计，即 Optimal Pseudo rewards，差别就在于，Optimal Pseudo rewards 单步最优即为整体最优，而 Bad Pseudo rewards 单步最优不一定是整体最优。我觉得这里可能借鉴了 CS 中贪心算法的一些局限性考察，有兴趣的可以了解一下。

或许还会有疑问，在各个条件下，如果被试稍微用一点心，就不会被限制得如此厉害。但是，不要忘了，每次移动后都有可能停止本 trail，这个限制因素我想是使得各实验设计能够发挥作用的重要前提，毕竟一旦需要额外 2 movements，您的飞机就只有不到 70% 的概率留在手中了。所以被试具有天然倾向于尽快到达 Smithsville（在补充材料中甚至被称为 wonderland🤣）

总结

继续研读了一下实验，我发现分享时可能存在一些误区。

1. 被试获得的实际奖励并不与线上显示的 points 挂钩，而是与其背后的 real rewards 挂钩。但是，在各实验设计中，简单走一走，会发现最大化 points 的路径与最大化 real rewards 的路径有一定的相关性，可并不是绝对！！！比如图 3 中，从 Brownsville 出发，经过 Clarksville 再到 Smithsville，所获得的 points 最多，但是经过 Clarksville 到 Bakersville 再到 Smithsville 所获得的 real rewards 最多。
2. 把一些实验设计的误区搞清楚了，Approximate Pseudo rewards 是通过距离表征，但是在路径选择上并不是取短为优（对于 points 也是如此）。Bad Pseudo rewards 反而是最简单的实验设计，只要靠近就有 +50 的 pseudo rewards。
3. Optimal Pseudo rewards 与 Approximate Pseudo rewards 方法的 pseudo rewards 计算方法还有待进一步研究，这点个人依旧觉得非常重要。

不过再强调一遍，实验一中分析的、被试获取奖励的指标是路径回归图 1 后，得到的 total real rewards，不是 points，不是 points，不是 points！