

# Minecraft を用いた状況付けられたタスク指向型対話データの収集

小河 晴菜<sup>1</sup> 徳永 健伸<sup>1</sup> 横野 光<sup>2</sup>

<sup>1</sup> 東京工業大学 情報理工学院 <sup>2</sup> 株式会社富士通研究所

{ogawa.h.ai@m, take@c}.titech.ac.jp yokono.hikaru@fujitsu.com

## 1 はじめに

自然言語処理では、大規模なデータを用いてモデルを学習する手法が主流となっており、対話処理の分野もこの例外ではない。雑談型の対話と比べてタスク指向型対話のデータ収集は、タスクに相当する既存のデータを見つけるのが難しいことから、タスクに応じて人間を用いたデータ収集が必要となる。本稿では、Minecraft (MC)<sup>1)</sup>上に構築したタスク指向型の対話データ収集基盤 [1](以下、DDCraft)を用いたデータ収集実験について報告する。MCをデータ収集基盤として使う利点は以下の2点である。

対話データの収集にクラウドソーシングを使う手法がある [2]。この場合、作業者の主な発話動機が対話目的の達成ではなく作業報酬となり、対話者が目的を達成するために行う発話とは動機が異なってしまう [3, 4]。この動機のずれが、タスク本来の対話からのずれとしてデータに反映される恐れがある。MCを土台にすることにより、タスクを容易にゲーム化でき、ゲーム参加者の発話の動機を対話目的に近づけることができる。

人間と物理的な場を共有するロボットや、道案内タスクなどでは、言語情報に加えて対話者が視覚的・空間的な状況を共有することが重要である。このような状況付けられたタスク指向型対話データの収集には、仮想世界を用いた例 [5, 6] や現実世界で行った例 [7, 8] がある。MCには仮想空間を操作するための基本的な枠組みが用意されているため、MCを土台にすることにより、様々なタスクの状況付けられた対話データの収集が容易になる。

## 2 関連研究

対象とするタスクにゲームの要素を導入する「ゲーミフィケーション」 [9] は、これまでもデー

タ収集で活用されている。画像のラベル付けにゲームを利用した ESP game [10] は、インターネット上に公開され1万人以上のプレイヤーを集め、アンテーションタスクにおけるゲーミフィケーションの有用性を示した。また、Vannellaら [3] はバリデーションタスクのための複数のゲームを提案し、インターネット上に公開したこれらのゲームの結果と、ゲームを用いない同様のタスクをクラウドソーシングの作業者に任せさせた結果を比べ、ゲーム形式で行った場合のほうが結果の質が高いことを示した。

対話収集への適用例としては、Manuvinakurikeら [11] が、クラウドソーシングによって話し言葉での対話を収集するブラウザゲームを作成した他、Asherら [12] が複数人で行うオンラインボードゲーム上にチャット機能を実装し、複数人の書き言葉での対話を収集している。

これらの研究はゲームを用いてタスク指向型対話を収集しているが、タスクが固定されており、他のタスクにゲームを流用することができない。一方、DDCraftはMC上における対話収集のための汎用的な機能を提供し、対話収集者はその上でタスクを設計するという点で異なる。これにより、対話収集者はMC本来の機能を活用しつつ実装の手間を削減することができる。

## 3 対話データ収集基盤

DDCraft [1] は、MCサーバ上の仮想世界で行われる対話について、ペアごとに、発話内容、発話時の発話者の位置情報と視界の情報、操作可能な物体の操作ログを収集する。サーバに接続した作業者らを自動でマッチングし、対話ペアを生成した後、ペアごとの個別の仮想世界を用意する機能を持つ。これにより、データ収集者は作業者のペア組みを調整する必要がない。さらに、DDCraftは仮想世界内の地図画像を出力するなど、ゲームを設計するための補

1) <https://www.minecraft.net/ja-jp>

助機能を提供する。これらの機能は MC の拡張機能として実装しており、MC にこれらを導入することで基盤を運用できる。

## 4 対話収集実験

DDCraft の基本的な機能を検証し、実際にどのようなデータが収集可能か、また、状況付けられた対話の特徴である参照表現の誘発が行えるかを確認するためにクラウドソーシングを使って参加者を募った。

### 4.1 対話タスク

対話データ収集には、DDCraft 上で行う対話タスクの例として我々が考案した Mansion Task を用いる。Mansion Task は協力的なタスク指向型対話を収集することを目的とした二人用のタスクで、Map Task [13] を元としている。このタスクでは、作業者はそれぞれ内容が一部異なる地図を渡される。地図には部屋の配置とゴール地点、更にゴールに到達するために行うべき複数のサブタスクの情報が記されている。作業者らの目的はこれらのサブタスクを達成し、ゴールに辿り着くことである。サブタスクを達成するための情報は二枚の地図に分割されて記載されるため、作業者らは対話によって情報を交換することになる。また、各サブタスクは二人が協力しないと達成できないよう設定されている。

本研究では、状況付けられた対話の特徴の一つである参照表現の誘発を目的としたサブタスクを設定した。Stoia ら [6] にならい、同一の物体を複数配置することで、特定の物体を表現しようとした作業者に参照表現を使用させることを狙う。実験に用いた Mansion Task の設定の詳細は付録を参照されたい。

### 4.2 実験方法

クラウドソーシングサービスの Amazon Mechanical Turk (MTurk)<sup>2)</sup> 上で、Mansion Task を用いた対話データ収集実験を行なった。MTurk では作業者が行うタスク (HIT) を登録することで仕事を依頼できる。また、その際に受諾条件を設定できる。作業者は自分が行いたい HIT を選択し、受諾する。

今回の実験では作業者同士がペアを組む必要がある。ペア組みには基盤のマッチング機能を用い、我々は個々のペア組みに関与せずに、サーバへの接続時間が合うよう HIT の説明内で大まかな接続推奨

時間帯を指定するにとどめ、自然にマッチングするようにした。

### 4.3 作業者の募集

我々はまず、募集人数を 20 人 (10 ペア分) として HIT を登録した。この試行では 5 人の作業者がサーバへ接続したが、接続時間がばらついておりペアが組まれなかった。また、HIT の登録後 10 分未満で受諾人数が募集人数に達した。受諾した作業者の中には、タスクを行わずに提出した作業や、実行せず放置した作業者が含まれている。短時間のうちにそういった作業で人数の枠が埋まってしまったことは、1 回目の試行が上手く行かなかった一つの原因だと考える。

1 回目の結果を踏まえて、設定を二点変更した。まず、接続時間のばらつきに対応するため、接続推奨時間帯を 15 枠から 3 枠に削減した。また、タスクを実行できない作業者が HIT を受諾するのを防ぐため、過去の成果の承認率と、MC のライセンスの有無を確認する設問への回答を受諾条件として追加した。その上で再度 HIT を登録したが、この 2 回目の試行ではサーバへの接続が一件のみだった。我々はこの原因が追加した設問にあると考えている。作業者が HIT の一覧を表示する際、自分が受諾条件を満たす HIT だけを抽出することができる。だが今回のような設問が設定されている場合、任意の作業者は初め条件を満たしていないため、抽出表示した作業者には HIT が表示されない。そのため、作業者数が大きく減ったのではないかと推測している。

設問を削除し、その分タスクを実行できない作業者が受諾しても募集人数が埋まらないよう、人数を 10 倍の 200 人に変更した。その上で 3 回目の HIT 登録を行なったところ、6 人の作業者が接続し、3 ペア分の対話を収集することに成功した。接続推奨時間帯を狭めたことで、作業者の接続時間を重ねることができたと考えている。また、受諾人数が募集人数に達するまで一日半程度の猶予があったため、タスクを行える作業者が受諾する余裕があったと考えられる。この試行では、基盤の設定不足が原因でサーバへの接続に失敗した作業者が 4 人いた。これは HIT 説明の設定手順が主に文章で構成されていたため、理解しづらかった可能性がある。

説明に図を追加した上で 4 回目の募集を行った。この試行では、13 人の作業者がサーバに接続した。うち 4 人は一度設定不足で接続に失敗したが、その

2) <https://www.mturk.com/>

後接続に成功している。これは、設定手順の説明のわかりやすさを改善したことが有効に働いたと考えている。しかしながら、ペアが成立したのは2組のみだった。接続ログを調べたところ、接続推奨時間帯外の接続が多かった(13人中7人)。これが、ペア成立数が伸びなかった原因であると考えている。

#### 4.4 募集における課題

4.3 節の結果から、DDCraft を用いた作業者の募集における二つの課題が示唆された。

一つ目は、DDCraft を用いるために、作業者が自ら拡張機能をインストールしなければならない点である。HIT を受諾したが放置した作業者が多かった点から、接続するための設定が手間であると思われる可能性がある。取得できる情報に制限がかかるものの、MC の拡張をサーバ側のみに留めるオプションを用意し、作業者が設定を行わずに接続できるようにすることで、この問題を緩和できると考える。

二つ目は、対話ペアを組むために、作業者が同時にサーバに接続している必要がある点である。DDCraft は作業者のマッチングを自動で行うため、データ収集者はペア組みを調整する必要がない。ただしその分、同じタイミングでサーバに接続した作業者がいなければ、作業者はタスクを行うことができない。これに対し、我々は接続推奨時間帯を設定したが、4 回目の例からも分かるように、作業者が必ずしもその時間帯に接続を行うわけではない。接続時間を指定するだけでなく、作業者のサーバ停留時間を延ばすことが重要であると考えられる。停留時間を延ばすための取り組みとして、我々は MC の特徴を活かし、作業者らが待機する仮想世界を探索して遊べるようなものに設定している。実験では、実際に遊んで待機する作業者が確認された。現在の仮想世界は単純なものであるため、ミニゲームの追加などより面白さを増やす改良を行うことで、作業者が同じ時間にサーバに接続している状況を作りやすくてできる可能性がある。

### 5 作業結果の考察

#### 5.1 作業者の動機づけ

表 1 にタスクを行った 10 人の作業者によるタスク後のアンケート結果を示す。アンケートは任意回答だったが、10 人全員から回答を得た。また、図 1 に自由記述コメントの抜粋を示す。アンケート

では、「もう一度遊びたいか」という問いに対して「Yes」の回答が 8 割を占めており、「報酬がなくても遊びたいか」についても、少し消極的な回答が増えるものの、「Rather Yes」以上の回答が 8 割を占めている。これらの回答やコメントの内容から、作業者が楽しんでタスクを行ったことが推測できる。このことから、ゲーム形式でタスクを行うことで、本研究が目的とする作業者の動機づけができていられる。また、全てのペアが最後までタスクを達成しており、DDCraft での対話収集では、基盤の設定やペア組みなどの壁があるものの、それらを乗り越えれば作業者が意欲的にタスクに取り組むことが示唆された。

表 1 タスク後の任意アンケート結果。RN は Rather No, Ne は Neither, RY は Rather Yes を表す。

質問	No	RN	Ne	RY	Yes
Would play again?	0	0	1	1	8
Would play with no reward?	0	1	1	5	3
Was setting up easy?	0	0	0	2	8

- Very fun and creative hit
- This was fun!!
- (前略) It was a great experience, and my partner and I did very well communicating!
- It was indeed a lot of fun, I got matched up with a chill player and we had a good time, thank's for this survey, I may end up playing this treasure hunt again.

図 1 作業者によるタスクへの自由記述コメント(抜粋)

#### 5.2 収集した対話データの特徴

表 2 に、収集した対話データの対話時間、発話数、参照表現および空間的な表現の出現回数を示す。ここで、参照表現・空間的な表現として「他との位置関係によって説明されているもの」「this, that などの指示語で説明されているもの」を数えた。一発話中に複数の表現が出現した場合、それぞれを一つとしてカウントする。また、対話 1~3 は収集実験の 3 回目の試行で、対話 4, 5 は 4 回目取得したものである。対話はすべて英語のチャットで行われた。

各ペアを比較すると、対話 1~3 と 4, 5 で発話数の合計が大きく違っている。しかし、3 回目と 4 回目の試行ではタスク内容を変えていないため、作業者の性格が原因で偶然この差異が現れたと考えている。また、対話 3 のペアでは参照表現が一度しか出現していない。作業者が入力したコメントによれば、このペアの主なコミュニケーションは MC 内のキャラクターの移動・腕の動きといった非言語情報

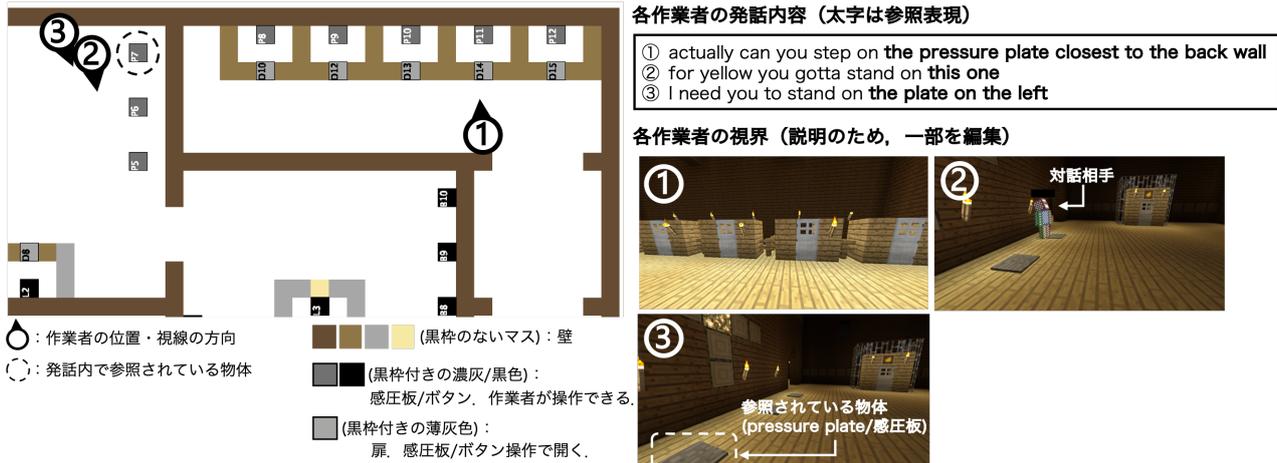


図 2 同一の対象に対する参照表現が現れた発話の内容、発話者の地図上の位置およびその視界. ここで、各発話は全て異なる対話のものである。また、地図は一部分のみを載せている。

で行われた、とあり、実際、収集した対話ログを見ると、このペアではタスクを達成するための操作に関する発言が2回しかされていない。

表 2 Mansion Task v2 を用いて収集したデータの情報。

対話	対話時間	発話数合計	参照表現個数
1	6分07秒	25	5
2	7分36秒	26	5
3	9分16秒	22	1
4	13分31秒	48	5
5	10分55秒	46	4
平均	9分29秒	33.4	4

図 2 に、ある同一の物体に対して用いられた参照表現を含む発話と、その際の周辺情報を示す。基盤は発話者の位置のみを記録するため、図には対話相手の位置情報は含まれない。三つの発話はそれぞれ別の対話であり、図内の破線で囲まれた物体(感圧板)を踏むよう発話者が相手に指示した際のものである。同じ対象を表すのに、三人の発話者が異なる表現を用いていることがわかる。それぞれの発話内容を見ると、対象から離れた位置で行われた1番の発話では建物内の位置を参照して対象を表現しているのに対し、対象と近い位置で起きた2, 3番の発話では発話者の位置を基準とした表現が用いられている。収集した対話データからは、このような作業者の位置による表現の違いや、周囲の情報がなければ意味を理解できない発話を確認できた。MCの仮想世界を利用することで、状況付けられた対話が行われたことがわかる。

また、図 2 の「作業者の視界」は、各作業者が実際に見ていた光景からゲームの UI を除いたものである。現実世界で作業者の視線や視界の情報を得る

ためには専用の機材が必要であるが、ゲームを用いれば、現実より単純化された視界ではあるものの、作業者の行動を妨げることなくこれらの情報を取得できる。これも MC を利用する利点の一つである。

以上の結果から、ゲームの世界であっても状況付けられた対話が収集できることが分かった。一方で、現状の基盤が記録していない、動作のような非言語情報や発話時の対話相手の位置情報が、より正確な状況の把握に必要であることが示唆された。それらの情報を取得できるよう基盤の改良を行うことで、より細かく作業者の行動を把握できると考えられる。

## 6 おわりに

本研究では、MC を用いた対話データ収集基盤 DDCraft 上で、クラウドソーシングによって状況付けられたタスク指向型対話の収集を行なった。その結果として、ゲーミフィケーションにより作業者が動機づけられることや、MC を用いることで状況付けられた対話を周辺情報と共に収集できることが確認できた。一方で、基盤を用いた対話収集における課題が確認された。今後は、大規模な対話データ収集のためにこれらの課題の有効な解決方法を検討する必要がある他、今回収集を行わなかった非言語情報を収集対象に含めるよう基盤の改良を行う。

## 謝辞

本研究は JSPS 科研費 JP19H04167 の助成を受けたものです。

## 参考文献

- [1] Haruna Ogawa, Hitoshi Nishikawa, Takenobu Tokunaga, and Hikaru Yokono. Gamification Platform for Collecting Task-oriented Dialogue Data. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 7084–7093, Marseille, France, May 2020. European Language Resources Association (ELRA).
- [2] Walter Lasecki, Ece Kamar, and Dan Bohus. Conversations in the crowd: Collecting data for task-oriented dialog learning. In *In Proceedings of the Human Computation Workshop on Scaling Speech and Language Understanding and Dialog through Crowdsourcing at HCOMP 2013.*, January 2013.
- [3] Daniele Vannella, David Jurgens, Daniele Scarfini, Domenico Toscani, and Roberto Navigli. Validating and extending semantic knowledge bases using video games with a purpose. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1294–1304, Baltimore, Maryland, June 2014. Association for Computational Linguistics.
- [4] Mikhail Burtsev and Varvara Logacheva. Conversational intelligence challenge: Accelerating research with crowd science and open source. *AI Magazine*, Vol. 41, No. 3, pp. 18–27, 2020.
- [5] Donna K. Byron and Eric Fosler-Lussier. The OSU quake 2004 corpus of two-party situated problem-solving dialogs. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, pp. 395–400, Genoa, Italy, May 2006. European Language Resources Association (ELRA).
- [6] Laura Stoia, Darla Magdalene Shockley, Donna K Byron, and Eric Fosler-Lussier. SCARE: A situated corpus with annotated Referring expressions. In *Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008*, pp. 650–653, 2008.
- [7] Susan E Brennan, Katharina S Schuhmann, and Karla M Batres. Entrainment on the Move and in the Lab: The Walking Around Corpus. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society, CogSci 2013*, pp. 1934–1939, 2013.
- [8] Kris Liu, Jean Fox Tree, and Marilyn Walker. Coordinating communication in the wild: The artwalk dialogue corpus of pedestrian navigation and mobile referential communication. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp. 3159–3166, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA).
- [9] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek '11*, p. 9–15, New York, NY, USA, 2011. Association for Computing Machinery.
- [10] Luis von Ahn and Laura Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '04*, pp. 319–326, New York, NY, USA, 2004. ACM.
- [11] Ramesh Manuvinakurike and David DeVault. Pair me up: A web framework for crowd-sourced spoken dialogue collection. In G.G. Lee, H.K. Kim, M. Jeong, and J.-H. Kim, editors, *Natural Language Dialog Systems and Intelligent Assistants*, pp. 189–201. Springer International Publishing, Cham, 2015.
- [12] Nicholas Asher, Julie Hunter, Mathieu Morey, Benamara Farah, and Stergos Afantenos. Discourse structure and dialogue acts in multiparty dialogue: the STAC corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp. 2721–2727, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA).
- [13] Anne H. Anderson, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson, and Regina Weinert. The HCRC map task corpus. *Language and Speech*, Vol. 34, No. 4, pp. 351–366, 1991.

## 付録：Mansion Task の詳細

本節では、本文中で触れた Mansion Task の設定について述べる。図 A1 に、作業員二人に分けて与えられる情報を纏めた Mansion Task の地図を示す。この地図は、DDCraft の地図画像生成機能が出力した画像に、矢印や説明書きを加えたものである。これらの説明は、作業員の発話に影響を与えないよう記号のみで記した。作業員に与えられる際は、一方の地図には図の“RED”と“YELLOW”に関する説明のみが、もう一方には“BLUE”と“GREEN”に関する説明のみが記載される。

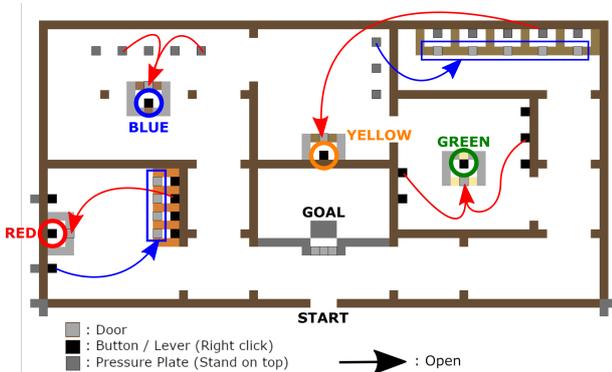


図 A1 Mansion Task の地図 (作業員二人分の情報を統合したもの)

このタスクの目的は、スタート地点の目の前にあるゴールに辿り着くことである。ただし、ゴールの前には扉があり、作業員らはその扉を開くために四つあるサブタスクを達成する必要がある。このサブタスクの達成方法は、すべて「閉ざされた扉を開け、その奥のレバーを起動する」というものである。扉の開け方はサブタスクごとに異なる。図 A2 にサブタスクの扉を示す。各サブタスクの扉の上には色のついたマークが記されており、これらはそれぞれ地図の RED, BLUE, YELLOW, GREEN に対応している。サブタスクの扉を開けるには、地図の情報に従って仮想世界内の操作可能な物体(ギミック)を操作する必要がある。Mansion Task のギミックは三種類あり、右クリックで起動できる「レバー」の他に、右クリックで起動できるが数秒後に起動状態が解除される「ボタン」、上に乗っている間起動できる「感圧板」が存在する。図 A3 にこれらのギミックを示す。

次に、各サブタスクの扉を開ける方法について述べる。RED は、片方の作業員がボタンを押すことで離れた位置にあるドアを開け、その間にもう一方の



図 A2 サブタスクの扉。扉の上に色のついたマークが示されている。図は RED



図 A3 ギミックの一覧。左から、ボタン、感圧板、レバー

作業員がドアの奥にあるボタンを押すことで達成できる。BLUE では、五つ並んだ感圧板のうち正しい二つを同時に踏むことでサブタスクを達成できる。YELLOW は、一方の作業員がある部屋の感圧板を踏んで別の部屋にあるドアを開け、もう一方がドアの奥にある感圧板を踏むことで達成できるサブタスクである。このサブタスクのみ、二人の作業員が別々の部屋で作業を行う。GREEN は、ある部屋の向かい合う壁にそれぞれ二つと三つ存在するボタンの中から、正しい二つを同時に押すことで達成できる。

RED と YELLOW は同様に「一方が扉を開け、もう一方がその奥のギミックを操作する」というサブタスクであるが、RED では作業員らが同じ部屋で作業を行い、相手の行動を見ることができのに対し、YELLOW ではそれぞれ別の部屋でギミックを操作する必要がある、自分の作業の結果や相手の動きが見えないという点で異なる。また、BLUE と GREEN はどちらもギミックを同時に操作するサブタスクであるが、感圧板の上に立ち続けていればよい BLUE に対し、GREEN はボタンを押した後の数秒間しかボタンが起動状態にならないため、タイミングを合わせなければならないという点で異なる。

これらのサブタスクの達成順序は問われないため、作業員らは自由な順序でそれぞれのサブタスクを行う。